

Understanding, Modeling and Detecting Brain Tumors : Graphical Models and Concurrent Segmentation/Registration methods

Sarah Parisot

▶ **To cite this version:**

Sarah Parisot. Understanding, Modeling and Detecting Brain Tumors : Graphical Models and Concurrent Segmentation/Registration methods. Other. Ecole Centrale Paris, 2013. English. NNT : 2013ECAP0064 . tel-00978520

HAL Id: tel-00978520

https://theses.hal.science/tel-00978520

Submitted on 14 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE CENTRALE PARIS

P H D T H E S I S

to obtain the title of

Doctor of Ecole Centrale Paris
Specialty : APPLIED MATHEMATICS

Defended by

Sarah PARISOT

Understanding, Modeling and Detecting Brain Tumors:

Graphical Models and Concurrent Segmentation/Registration Methods

prepared at Ecole Centrale Paris, Center for Visual Computing

defended on November 18, 2013

Jury :

Reviewers :	Christos DAVATZIKOS	-	University of Pennsylvania
	Alison NOBLE	-	University of Oxford
Advisor :	Nikos PARAGIOS	-	Ecole Centrale Paris
Examiners :	Nicholas AYACHE	-	INRIA - Sophia-Antipolis
	Stéphane CHEMOUNY	-	Intrasense SAS
	Emmanuel MANDONNET	-	Hôpital de Lariboisière
	William WELLS III	-	Harvard Medical School

Acknowledgments

First of all, I would like to express my sincere gratitude towards my advisor Nikos Paragios for his continuous support and advice during my whole PhD. It has been a long journey, and I certainly wouldn't be where I am now if it wasn't for him. Then, I would like to thank Intrasense and more specifically its CEO Stéphane Chemouny, as well as Hugues Duffau for trusting me with this project and for the freedom I was given during this thesis. I will always remember the incredible opportunity I was given to observe a brain surgery at the beginning of my thesis, which was a truly fascinating experience.

Second, I would like to express my deepest gratitude towards all the members of my thesis committee: the reviewers Alison Noble and Christos Davatzikos, the chairman Nicholas Ayache and the examiners William Wells III and Emmanuel Mandonnet, for taking time to read and evaluate my work and their insightful comments and questions.

I would like to thank William Wells for hosting me in the Surgical Planning Laboratory and the three great months I spent there. I was an incredibly interesting and fruitful experience that completely renewed my motivation for my work. I would also like to thank all the members of the lab for their warm welcome and the fun times we spent together.

I thank Amélie Darlix for the enormous work she carried out constructing the DLGG database and her help and advice on the clinical aspect of my work.

I would like to acknowledge Bo (aka Crazy Bobo) who has been a wonderful (the best!) colleague and friend throughout those years. Thank you for all the experiences we shared together, I am very glad that I had you by my side during the tough (such as our first MICCAI deadline) and the happy times. I would also like to thank Blandine for being such a great friend, I am thankful that we started our crazy PhD adventure almost simultaneously. To a larger extend, I thank all my great labmates for making this lab such a better place (and for eating my million cakes): Katerina, Aris, Loïc, Olivier, Chaohui, Panos, Pierre-Yves, Nicolas, Vivien, Sylvain, Fabrice, Haithem, the Stavroses, Nacho, Enzo, Eduard, Puneet, Krishna, Siddharta, Dimitris, Grigoris, Evgenios, Stefan, Mateuzs, Matthew, Iasonas and Pawan. Thank you for the countless times you have helped me with work, the Chinese and belated Greek lessons, the tea/coffee breaks, the sweets from all over the world, the dinner parties...

I would like to thank Sylvie and Annie from MAS laboratory, as well as Carine and Natalia from CVN for their invaluable help and support as well as the great discussions we had together. I would also like to acknowledge my former ECP teachers Moncef Stambouli and Florence Mayo-

Quenette with whom I was always happy to cross path and have a chat.

Outside the lab, I would like to thank my dear friends who have supported me during all those years and made sure I had a social life, even during the dreaded thesis writing. Life would have been much harder without my precious high school friends Juliette (one of my tiniest yet greatest friend), Nicolas (my photo wander partner), Olivier (my favorite Swiss) and Pauline (when we manage to see each other); my Intrasense penpal Guillaume; and of course, my wonderful ECP family "Centraliénés" Alexandra, Aline, Amélie, Anne-Gabrielle, Céline, Eric, Gabrielle, Guillemette, Philippe and Victor.

Finally, I dedicate this thesis to my parents and sisters, whose unconditional love and support got me through the hardest moments and with whom I was glad to share the happy ones. I feel incredibly lucky to have such a great family and I love you all very deeply.

Abstract

The main objective of this thesis is the automatic modeling, understanding and segmentation of diffusively infiltrative tumors known as Diffuse Low-Grade Gliomas. Two approaches exploiting anatomical and spatial prior knowledge have been proposed.

We first present the construction of a tumor specific probabilistic atlas describing the tumors' preferential locations in the brain. The proposed atlas constitutes an excellent tool for the study of the mechanisms behind the genesis of the tumors and provides strong spatial cues on where they are expected to appear. The latter characteristic is exploited in a Markov Random Field based segmentation method where the atlas guides the segmentation process as well as characterizes the tumor's preferential location.

Second, we introduce a concurrent tumor segmentation and registration with missing correspondences method. The anatomical knowledge introduced by the registration process increases the segmentation quality, while progressively acknowledging the presence of the tumor ensures that the registration is not violated by the missing correspondences without the introduction of a bias. The method is designed as a hierarchical grid-based Markov Random Field model where the segmentation and registration parameters are estimated simultaneously on the grid's control point. The last contribution of this thesis is an uncertainty-driven adaptive sampling approach for such grid-based models in order to ensure precision and accuracy while maintaining robustness and computational efficiency.

The potentials of both methods have been demonstrated on a large data-set of heterogeneous -in appearance, size and shape- Diffuse Low-Grade Gliomas. The proposed methods go beyond the scope of the presented clinical context due to their strong modularity and could easily be adapted to other clinical or computer vision problems.

Keywords: *Markov Random Fields, Registration, Segmentation, Brain Atlas, Diffuse Low-Grade Gliomas*

Résumé

L'objectif principal de cette thèse est la modélisation, compréhension et segmentation automatique de tumeurs diffuses et infiltrantes appelées Gliomes Diffus de Bas Grade. Deux approches exploitant des connaissances a priori de l'ordre spatial et anatomique ont été proposées. Dans un premier temps, la construction d'un atlas probabiliste qui illustre les positions préférentielles des tumeurs dans le cerveau est présentée. Cet atlas représente un excellent outil pour l'étude des mécanismes associés à la genèse des tumeurs et fournit des indications sur la position probable des tumeurs. Cette information est exploitée dans une méthode de segmentation basée sur des champs de Markov aléatoires, dans laquelle l'atlas guide la segmentation et caractérise la position préférentielle de la tumeur.

Dans un second temps, nous présentons une méthode pour la segmentation de tumeur et le recalage avec absence de correspondances simultanés. Le recalage introduit des informations anatomiques qui améliorent les résultats de segmentation tandis que la détection progressive de la tumeur permet de surmonter l'absence de correspondances sans l'introduction d'un à priori. La méthode est modélisée comme un champ de Markov aléatoire hiérarchique et à base de grille sur laquelle les paramètres de segmentation et recalage sont estimés simultanément. Notre dernière contribution est une méthode d'échantillonnage adaptatif guidé par les incertitudes pour de tels modèles discrets. Ceci permet d'avoir une grande précision tout en maintenant la robustesse et rapidité de la méthode. Le potentiel des deux méthodes est démontré sur de grandes bases de données de gliomes diffus de bas grade hétérogènes (en taille, forme et apparence). De par leur modularité, les méthodes proposées ne se limitent pas au contexte clinique présenté et pourraient facilement être adaptées à d'autres problèmes cliniques ou de vision par ordinateur.

Keywords: *Champs de Markov Aléatoires, Recalage, Segmentation, Atlas du Cerveau, Gliomes Diffus de Bas Grade*

Contents

1 Introduction **15**
- 1.1 Context and Motivation . . . 15
  - 1.1.1 Diffuse Low-Grade Gliomas: Introduction . . . 16
  - 1.1.2 Therapy and Functional Reshaping . . . 16
  - 1.1.3 Follow-up: Tumor Segmentation . . . 17
  - 1.1.4 Low-Grade Gliomas' Localization in the Brain . . . 19
- 1.2 Thesis Overview . . . 19

2 Brain Tumor Segmentation **21**
- 2.1 Introduction . . . 21
- 2.2 Brain Imaging . . . 22
  - 2.2.1 Imaging modalities . . . 22
  - 2.2.2 Magnetic Resonance Imaging . . . 24
- 2.3 Surface Based Segmentation: Deformable models . . . 28
- 2.4 Region-Based Segmentation . . . 31
  - 2.4.1 Classification methods . . . 31
  - 2.4.2 Graph based Segmentation . . . 35
- 2.5 Atlas based Segmentation . . . 43
  - 2.5.1 Medical Image Registration . . . 43
  - 2.5.2 Brain Atlases . . . 48
  - 2.5.3 Atlas based Tumor Segmentation . . . 50
- 2.6 Summary . . . 53

3 A Probabilistic Atlas of Diffuse WHO Grade II Gliomas' Preferential Locations **55**
- 3.1 Introduction . . . 55
- 3.2 Spatial Position Mapping of Diffuse Low-Grade Gliomas . . . 56
  - 3.2.1 Graph Construction . . . 56
  - 3.2.2 Graph Clustering . . . 58
  - 3.2.3 Clustering Selection . . . 61
- 3.3 Tumor Characterization and Segmentation with Spatial Position Priors . . . 64

3.3.1 Tumor Detection . . . 64
3.3.2 Coupled Markov Random Field Model . . . 68
3.4 Experimental Validation . . . 72
3.4.1 Data-set and Preprocessing . . . 72
3.4.2 Sparse Graph Validation and Robustness . . . 73
3.4.3 Segmentation Results . . . 77
3.5 Discussion . . . 80

4 Joint Tumor Segmentation and Dense Deformable Registration 85
4.1 Introduction . . . 85
4.2 Related work . . . 86
4.2.1 Registration with Missing Correspondences . . . 87
4.2.2 Joint Registration and Segmentation . . . 91
4.3 Concurrent Deformable Registration and Tumor Segmentation . . . 93
4.3.1 Discrete MRF Model for Deformable Registration . . . 94
4.3.2 Coupled Tumor Segmentation and Registration . . . 95
4.4 Uncertainty Driven Resampling . . . 99
4.4.1 Local Registration and Segmentation Uncertainties . . . 99
4.4.2 Min Marginals Computation using Dynamic Graph Cuts . . . 100
4.4.3 Content Driven Adaptive Resampling . . . 101
4.5 Experimental Validation . . . 104
4.5.1 Data-Set Preprocessing . . . 104
4.5.2 Implementation . . . 105
4.5.3 Results . . . 107
4.6 Discussion . . . 108

5 Conclusion 115
5.1 Contributions . . . 115
5.2 Future Work . . . 116

A Concurrent Segmentation and Registration Application: Registration of Preoperative and Intraoperative Images with Tumor Resection 121
A.1 Introduction . . . 121
A.2 Related work . . . 122
A.3 Probabilistic Formulation . . . 124
A.4 Likelihood term . . . 126
A.4.1 Resection Cavity Detection . . . 126
A.4.2 Background area . . . 129
A.5 Experiments . . . 130
A.5.1 Implementation Details . . . 130
A.5.2 Results . . . 131

A.6 Discussion 131

List of Figures

1.1 Preoperative functional mapping. The functional areas are reorganized around the tumor's boundaries. [Duffau 2006] . . . 17
1.2 Intraoperative brain mapping. (a) Preoperative MRI, (b) Intraoperative Mapping. Each tag correspond to a functional area that is identified through cortical stimulation, (c) Postoperative MRI. Images extracted from [Duffau 2006] . . . 18
1.3 Modeling of the LGG's growth rate. (a) Linear growth rates of 27 patients, displaying a mean growth rate of 4mm/year. (b) Growth rate evolution of a patient after anaplastic transformation. Figures extracted from [Mandonnet 2003] and [Mandonnet 2008] respectively. . . . 19

2.1 Examples of different DLGG appearances on FLAIR T2 MRIs. . . . 22
2.2 Impact of an magnetic field B_0 on the spins' magnetization. (a) Random spins orientation, (b) alignment with the external field, (c) spin precession around B_0. Images taken from [Puddephat 2002] . . . 25
2.3 Impact of TE and TR on the NMR signal [Buxton 2009]. (a) NMR signal, amplitude dependent on T2, (b) Impact of TR on the initial amplitude of the signal, (c) longitudinal magnetization dependent on TR and T1. . . . 26
2.4 Contrast enhancement of High grade gliomas. (a) T1 weighted image, (b) contrast enhanced T1 weighted image. . . . 27
2.5 Images sequences for the visualization of a Low grade glioma. (a) T1, (b) T1 contrast enhanced (no enhancement of the tumor), (c) T2, (d) FLAIR images. . . . 28
2.6 Level set of an implicit function u and corresponding curve C [Leventon 2000]. . . . 29
2.7 Figure extracted from [Droske 2001]. Adaptive grid sampling, progressively constructed as the front progresses. The colors indicates the arrival time of the propatation (from blue (early) to red) . . . 30
2.8 Illustration of the kNN clustering approach. We aim at classifying the black sample as either a blue square or red circle. Setting k=5 selects the circle class, while k=8 (dashed circle) selects the other class. . . . 33
2.9 Illustration of the SVM approach. . . . 34
2.10 Illustration of the SWA algorithm extracted from [Corso 2008]. . . . 37

2.11 Evolution of the joint histogram between an image and iself as one image is progressively rotated (0,2, 5 and 10 degrees). Below the histogram are the joint entropy values. Example taken from [Pluim 2003] 45
2.12 Different types of transformation models. (a) original image, (b) rigid transformation, (c) affine transformation, (d) non-linear transformation. 46
2.13 Schematic illustrating the AC-PC line and axes of the Talairach space. Example taken from http://airto.ccn.ucla.edu/. 49
2.14 MNI-ICBM-152 probabilistic atlas. From left to right: T1, T2 and Proton Density weighted images ; Gray Matter, White Matter and CSF probability maps. Figure extracted from [Fonov 2011] 50
2.15 Modified MNI-305 atlas for integration of tumor and edema probability [Prastawa 2003]. Top row: healthy probability (White Matter, Grey Matter, CSF). Bottom row: tumor(left) and edema (right) probabilities. 52

3.1 Illustration of the block matching process. The floating image is on the left and target on the right. Figure extracted from [Ourselin 2000] 58
3.2 Visualization of the complete graph superimposed to the mean registered image. For visibility reasons, edges of weight larger than 35 are not shown. Short distances appear red and thick while large distance values are thin and blue. 59
3.3 Probability maps of three different clusters superimposed to the mean registered image. 65
3.4 Visual examples of features used for boosting training: (a) median filter, (b) standard deviation, (c) entropy, (d) symmetry, (e,f) Gabor features. 68
3.5 Image courtesy of [Komodakis 2008b]. (a) Primal-Dual principle: the optimal cost always lies between $\mathbf{b}^T\mathbf{y}$ and $\mathbf{c}^T\mathbf{x}$ for any pair $(\mathbf{x}, \mathbf{y})$ of feasible solutions. If both values are close, the approximated solution to the primal problem is close to the optimal solution too. (b) Primal-Dual schema: iteratively generate pairs $(\mathbf{x}^k, \mathbf{y}^k)$ of feasible solution until they are close enough (their ratio is below $f*$) to obtain an $f*$-approximation of the solution 71
3.6 Distribution of tumors sizes in the data set (number of tumors vs tumor size in cm^3) 73
3.7 Visualization of the complete clustered graph superimposed to the mean registered image. The numbers correspond to the number of nodes in each cluster. 74
3.8 Schematics of the brain structure. 75
3.9 Cluster validity indices with respect to the value of α (a,b,c,d) and the number of clusters (e). Dunn index (a), Davies Bouldin index (b), Silhouette index (c) and combined indices (d,e). 76
3.10 Errorbars of the indices computed for the cross validation experiments with respect to the value of α (a) and the number of clusters (b). 77
3.11 Illustration of the graph matching results. (a) Complete match, (b,c) Partial match. Positive matches correspond to blue edges and mismatched samples to red edges. . 78

3.12 Boxplots of the Dice score (a), True Positive rate (b), False positive rate (c), and MAD score (d) between the automatic and manual tumor segmentation for the three different methods. 79
3.13 3D visualization of an example segmentation result. (a) manual segmentation, (b) automatic segmentation 80
3.14 Positions of the cluster centers on the ICBM atlas after registration. A strong symmetry can be observed between clusters in columns (a) and (b) of the same image row. 82
3.15 Visual Segmentation results. (a) boosting score, (b) Boosting classification (thresholding), (c) Pairwise MRF, (d) MRF with spatial prior. 83

4.1 Overview of the Joint Registration and Segmentation method 86
4.2 Summary of different steps of the ORBIT algorithm. Figure extracted from [Zacharaki 2008]. 90
4.3 Modified ICBM probabilistic atlas by the tumor growth model. From left to right: tumor, edema, CSF, gray matter and white matter probabilities. Figure extracted from [Gooya 2012] 93
4.4 Registration uncertainty and displacement set resampling for one control point: (a) Min marginal values per displacement label (blue: low, red: high energy) associated covariance matrix centered at the optimal label, (b) Min marginals visualization on a 2D slice, (c) Original isotropic displacement set and (d) Uncertainty driven displacement set, following the brain boundaries. 102
4.5 Visual representation of the grid refinement from level j (left) to level j+1 (right). Grid resampling: the nodes that have direct correspondences appear in white, and the new nodes and edges are red. The edges connecting the 2 grids represent the nodes' neighborhood. The grid is shown in 2D for increased visibility. 103
4.6 The 4 levels of the incremental approach. 106
4.7 The two displacement sampling strategies, centered around the node's initial position: (a) Dense displacement sampling, (b) Sparse displacement sampling. The red dots correspond to the possible displacements. 106
4.8 Mean percentage of activated nodes per resolution level. 107
4.9 Visual examples of the activated nodes for the 4 levels of the incremental approach. The nodes are superimposed to the target image. 110
4.10 Quantitative Segmentation Results: Error bars (mean and standard deviation) of the Dice score, False Positive (FP) and True Positive (TP) rates and MAD score for the joint framework with low (JSRLow) and high resolution (JSRHigh), the MAP-MRF segmentation framework (SegMRF) and the uncertainty based approach (Ucy). 111
4.11 Quantitative Registration Results: Error bar graphs of the Dice, True Positives (TP), False Positives (TP) and MAD scores obtained for the joint framework with low (JSRLow) and high resolution (JSRHigh), the individual registration framework with masked pathology (RegMask) and the uncertainty based approach (Ucy). 111

4.12 Visual examples of Registration results. (a) Target image, deformed image obtained using (b) the individual framework with masked pathology, (c) the Joint framework, high resolution, (d) the Joint framework, with adaptive sampling. . . . 112
4.13 Visual examples of Segmentation results. (a) Boosting Score, (b) MAP-MRF framework, (c) Joint framework, high resolution, (d) Joint framework, with adaptive sampling. Automatic segmentations (blue) are compared to the manual segmentation (red) 113

A.1 Gaussian Mixture Model of the resection cavity intensity. (a) Intensity histogram, (b) Modeled intensity distribution, (c) Gaussians constituting the GMM. 128
A.2 Log probabilities associated to the resection cavity coupled with the spatial prior. . 129
A.3 Difference Image after affine registration (a) and after non-rigid registration (b). . . 132
A.4 Visual registration and segmentation results. (a) Preoperative image after affine registration, (b) non-rigidly deformed preoperative image, (c) intraoperative image segmented. Resection cavity segmentation: automatic (blue) and manual (red); Residual tumor segmentation: manual (magenta) and registered preoperative tumor (yellow). 133

List of Tables

3.1 Anatomical localization of the different clusters 75
3.2 Mean and median (in parentheses) values of the different scores and methods for tumor segmentation. 80

Chapter 1

Introduction

This thesis introduces methods for the study and monitoring of brain tumors and in particular the *diffuse WHO grade II gliomas* or *Diffuse Low-Grade Gliomas* (DLGG). This is carried out by focusing on two challenging and important problems in medical imaging: *Image Segmentation* and *Image Registration*. The segmentation task consists of contouring an object of interest in an image in order to provide its exact position. The task becomes increasingly difficult as the object of interest's appearance gets more heterogeneous and has irregular boundaries. The complemenary task of image registration aims at aligning two or more images by finding corresponding structures between them. The difficulty resides in the relative appearance of the images and fuzziness of the structures' edges.

This chapter features a presentation of the clinical context associated to DLGGs and the motivation and main objectives of this thesis, followed by an overview of the thesis organization.

1.1 Context and Motivation

Primary brain tumors constitute a heterogeneous set of tumors associated with variable behaviors, symptoms, origins and malignancy. They are characterized by the type of cell from which they arise and are dominated by gliomas, arising from glial cells. Gliomas are classified by the World Heatlh Organisation (WHO) into 4 grades of malignancy based on their histological properties [Louis 2007].

Grade I gliomas (mainly pilocytic astrocytomas) are benign tumors, as they are well delineated and non infiltrative. They have an excellent prognosis after complete surgical resection. Diffuse grade II gliomas are slow-growing infiltrative tumors with continuous growth [Mandonnet 2003]. They inevitably progress into a malignant tumor. Grade III gliomas (or anaplastic gliomas) are fast growing malignant tumors while grade IV gliomas (glioblastomas multiforme) are the most aggressive tumors with a low treatment success rate. They are unfortunately the most common type of glioma.

1.1.1 Diffuse Low-Grade Gliomas: Introduction

DLGGs are diffusely infiltrative primary brain tumors. They represent about 15 % of primary brain tumors [Rigau 2011, Dolecek 2012], and affect younger patients (slightly more males than females) than the most common glioblastomas (mean age about 38 years old) [Capelle 2013]. They are classified into 3 histological subtypes according to their predominant cell type: astrocytomas, oligodendrogliomas and oligoastrocytomas.

Most of the time, DLGGs are revealed by seizures that occur in about 80 % of the cases. Some mild symtoms such as headaches, nausea, speech disturbance can appear depending on the size and location of the tumor. This absence of symptoms rendered the treatment of the tumors very controversial as a surgical tumor resection could cause functional deficits that could be more important than the ones produced by the tumor itself. In the past years, a "wait and see" policy was predominantly considered, where the tumors' growth was simply followed by successive MRIs [Whittle 2010]. It has been observed however, that during their natural course, the tumors inevitably progress into higher grade fast growing tumors (anaplastic transformation into WHO grade III gliomas) which lead to strong neurological deficits and ultimately to the patient's death (median survival is about 10 years). Therefore, a LGG has to be monitored closely in order to adapt the treatment to its growth rate and corresponding prognosis.

1.1.2 Therapy and Functional Reshaping

Nowadays, a consensus has been reached that maximal safe surgical resection is the first therapeutic option for DLGG [Soffietti 2010]. Indeed, several studies reported an improvement in the prognosis of progression-free and overall survival of patients after surgery [Sanai 2008, Capelle 2013, Jakola 2012]. Tumors localized in or near an important functional area used to be considered inoperable. However, it has been shown that some patients could recover without permanent deficit after damages of eloquent areas [Duffau 2003, Duffau 2006]. Other studies show that functional recovery is better in cases of slow growing lesions (like the LGG) than in cases of fast growing lesions (such as strokes) [Desmurget 2007]. It is inferred that this is due to a functional reshaping of the brain induced by the slow growing lesions, which has no time to occur when the damage is rapid. This phenomenon is also known as brain plasticity. Different patterns of functional organization with respect to the tumor can be considered. Firstly, the function can persist within the tumor (36% of patients present intra tumoral activity in [Desmurget 2007]). In that case, it is quasi impossible to completely remove the tumor without sequelae. Another possibility is the redistribution of eloquent areas immediately around the tumor. A near total resection is possible then, but yields transient deficit after the surgery. Thirdly, regions implicated in the functional network in the same hemisphere but remote to the damaged area can be recruited. Finally, the other hemisphere can be solicited. This reshaping would explain the absence of symptoms even when the tumor is located in eloquent areas.

Functional imaging is routinely used for surgery planning in order to evaluate the relationship between the tumor's location and the brain's functional areas and determine the extent of resection

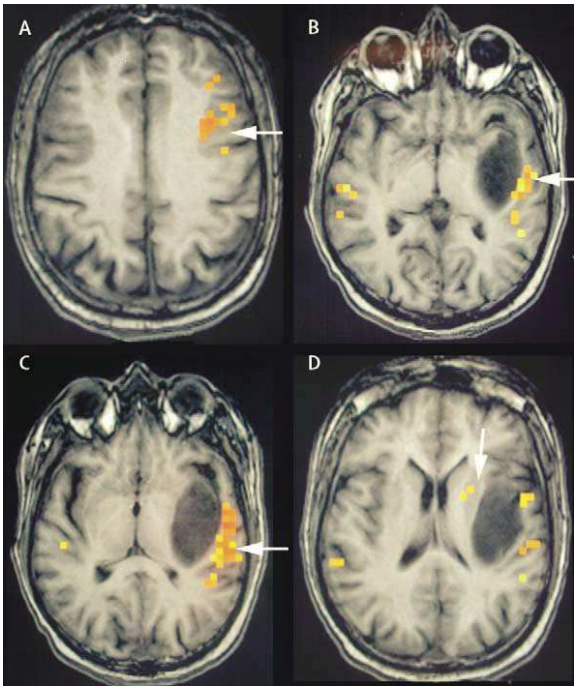


Figure 1.1: *Preoperative functional mapping. The functional areas are reorganized around the tumor's boundaries. [Duffau 2006]*

possible. However, it is not sufficient in order to evaluate if the eloquent area is essential or compensable, while precision decreases around the lesion [Holodny 2000]. A solution is to perform brain mapping during the tumor surgery. Direct stimuli are sent on the surface of the brain, which enables the mapping of the motor functions and the cognitive areas (if the patient is awake). If any functional disturbance is noticed, the resection is stopped immediately in that area. The tumor is then removed according to the functional boundaries. Radiotherapy and chemotherapy can be used after surgery for patients with large or fast growing residual tumors.

1.1.3 Follow-up: Tumor Segmentation

In addition to surgery planning, medical imaging plays a key role in DLGG therapy planning and evaluation. Since the tumors are mostly asymptomatic, their growth is monitored by serial MRI exams (usually every 3 months). The best way to follow the tumor's progression is to observe its mean diameter that is expected to follow a linear law [Mandonnet 2008]. Currently, two methods exist for evaluating the mean diameter. The simplest approach consists in manually measuring the tumor's three biggest diameters (in the sagittal, coronal and axial planes) and approximating the tumor's volume as an ellipsoid. The alternative is to compute the tumor's volume using manual segmentation. The mean diameter is finally obtained from the volume as

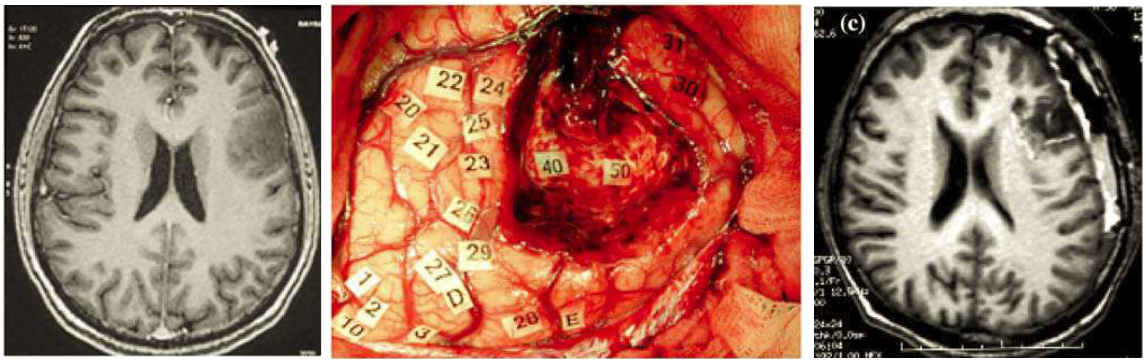
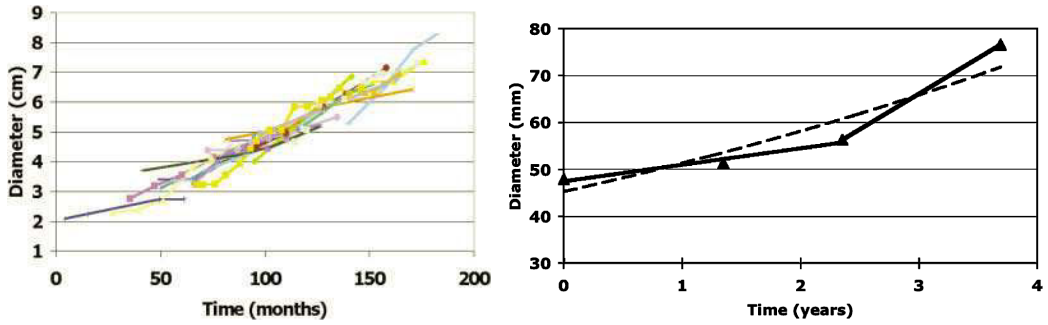


Figure 1.2: *Intraoperative brain mapping. (a) Preoperative MRI, (b) Intraoperative Mapping. Each tag correspond to a functional area that is identified through cortical stimulation, (c) Postoperative MRI. Images extracted from [Duffau 2006]*

$D_{mean} = (2V)^{1/3}$ [Mandonnet 2003, Mandonnet 2008]. Despite being fast and convenient, the three diameters method often overestimates the volume of the tumor, especially in cases of irregular contours (which is often the case for low grade gliomas due to their infiltration properties). It can yield inexact results (especially during follow ups after resection, when the shape of the tumor is very different from an ellipse). Therefore, the 3D manual segmentation is currently the gold standard to evaluate the mean diameter. It is, however, tedious and time consuming and suffers from a high intra and inter operator variability [Kaus 2001].

Typically, the mean diameter progresses according to a linear law. The slope is constant (mean value 4 mm/year) until anaplastic transformation, where there is an abrupt change of speed (average of 8 mm/year). The evaluation of the growth rate is essential for LGG management as it is directly associated to the patient's prognosis and risk of anaplastic transformation. Pallud et al. [Pallud 2012] proposed to systematically evaluate the growth rate and adapt the treatment accordingly. Such evaluation was based on measuring the growth rate between two temporal MRIs separated by 4 months. A growth rate higher than 2mm/year triggers a surgical treatment. An alternative was to use Chang's score [Chang 2009] that is based on 4 risks factors: 1) tumor located in an eloquent area, 2) Karnosky Performance Status ≤ 80 (measure of functional impairment, a score higher than 80 implies no impairment), 3) age > 50 years old, 4) tumor's maximum diameter $> 4cm$. If the patient obtains a score of $3-4$ a second MRI is performed after 6 weeks. A mean tumor diameter growth larger than 1 mm suggests a growth rate greater than 8.7 mm/year implicating a behavior similar to the one of a high grade glioma. For a score lower than 3, the second MRI is to be performed after 3 months and the mean diameter is measured. The treatment is adapted accordingly.

This strong dependence on the growth diameter as well as the limitations of manual segmentation and ellipsoid approximation methods highlight the need for a fast and robust (*i.e.* independent of the observer) automatic segmentation method, that is the main objective of this thesis.



(a) (b)

Figure 1.3: *Modeling of the LGG's growth rate. (a) Linear growth rates of 27 patients, displaying a mean growth rate of 4mm/year. (b) Growth rate evolution of a patient after anaplastic transformation. Figures extracted from [Mandonnet 2003] and [Mandonnet 2008] respectively.*

1.1.4 Low-Grade Gliomas' Localization in the Brain

As stated earlier, the treatment of low-grade gliomas strongly depends on their localization. It was inferred in [Duffau 2004] that they tend to appear in preferential locations that correspond essentially to functional areas with a predominance in the sensory-motor area and the insula (that studies tend to associate with language functions [Ackermann 2004]). This observation is the starting point of the second objective of this thesis, which is the construction of a statistical atlas mapping the low-grade gliomas preferential locations. Such an atlas offers diverse possibilities of understanding the tumors better. First, the position of the tumors could be correlated with extrinsic factors in order to identify risk factors associated to the DLGG. Second, comparison with functional areas would allow to study the mechanisms of brain plasticity associated to the slow growing tumors. Eventually, we will show that mapping the tumors in the brain provides a useful tool for tumor detection and segmentation.

1.2 Thesis Overview

This thesis is divided into 3 main chapters. Chapter 2 presents an overview of the different brain tumor segmentation methods existing in the literature, coupled with an introduction to the image registration problem for atlas-based segmentation.

In chapter 3, the construction of a low-grade gliomas probabilistic atlas is presented. The atlas enables to identify the tumors' preferential locations in the brain while describing their spatial extension in each preferential location. The second part of this chapter presents an application of the atlas for brain tumor segmentation, using it as a spatial prior on where the tumor is expected to be.

Chapter 4 introduces a concurrent registration and segmentation method that simultaneously

segments the brain tumor, and registers a healthy subject to the diseased subject with a tumor while taking into account the missing correspondences area represented by the tumor. The chapter presents an alternative method for tumor segmentation that exploits anatomical cues via the registration process, as well as a more precise registration scheme for a future instance of the probabilistic atlas. An extension of the method for the registration of preoperative and intraoperative MRIs during a brain tumor surgical resection is be presented in appendix A.

Chapter 5 concludes the thesis with a discussion on the presented work and its implications, and suggests future research directions with respect to the presented methods.

Chapter 2

Brain Tumor Segmentation

2.1 Introduction

Image segmentation is one of the most important and active research area in the medical imaging domain. It can be defined as the delineation of one or several structures of interest within the image. Automated methods are sought in order to avoid the time consuming burden of manually contouring the structures. The problem is particularly difficult in the context of brain tumors. Indeed, most tumors have heterogeneous appearances and their intensity range overlap with the healthy tissues'. The presence of a necrotic core is frequent (especially for glioblastomas, but it also occurs for DLGGs) resulting on a strong contrast with the "active" tumor. Prior information regarding the shape of the tumor cannot be used as they have variable sizes and shapes. DLGGs in particular, have very fuzzy and irregular boundaries due to their infiltrative nature. Edema (swelling of brain tissue around the tumor) and mass effect (tissue displacement induced by the tumor) are quite uncommon due to the slow-growing nature of the DLGGs [Sanai 2011]. In this context, the simplest segmentation methods such as thresholding or region growing are insufficient [Gibbs 1996]. Despite extensive and promising work in the tumor segmentation field, obtaining accurate and reliable segmentations of brain tumors remains a difficult task.

Segmentation methods can be grouped in two categories: surface and region-based approaches. The objective of surface based methods is to find the organ or tumor's boundary by propagation a curve/surface with a flow that is determined according to curvature and image constraints (generally the image gradient). Snakes and level sets are typically used in this context. The former defines the object's boundary explicitly as a parametric curve, while the latter defines the contour via an implicit function allowing for more complex geometries and topological changes.

Region based methods consider the segmentation problem from a different angle. Here, the goal is to identify all voxels belonging to the object and separate them from the rest of the image. Early work relied on fuzzy clustering to regroup similar voxels but quickly showed limitations. Supervised statistical pattern classification techniques have been the basis of the majority of recent region based tumor segmentation methods. The voxels are separated by a classification score or

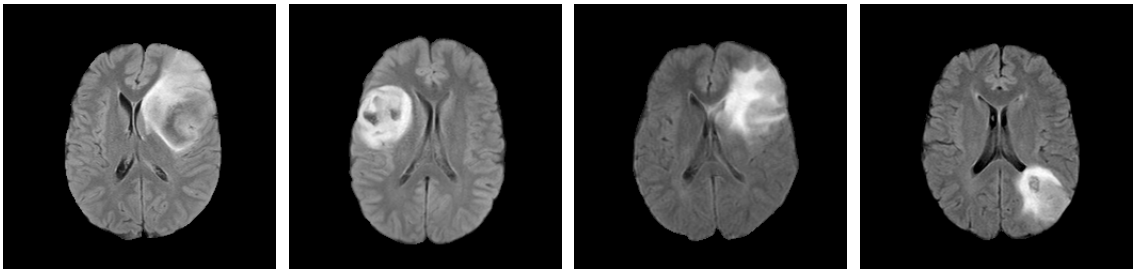


Figure 2.1: *Examples of different DLGG appearances on FLAIR T2 MRIs.*

probability, the ultimate goal being to label each pixel of the image as tumor or background. The tumor is frequently separated into active tumor, necrosis and an additional class is introduced in the context of fast growing tumors for the induced edema around the tumor. The obtained segmentations are refined by the use of graph based methods that model spatial dependencies and prior anatomical knowledge.

The remainder of this chapter is organized as follows: section 2.2 consists of an overview of the different medical imaging modalities used for brain imaging on which the segmentation algorithms are applied, followed by a review the different existing tumor segmentation methods (including, but not limited to DLGG segmentation methods) in the remaining sections. Section 2.3 will be dedicated to surface based methods while Section 2.4 will present the region based methods for tumor segmentation. Section 2.5 will focus on Atlas based tumor segmentation, a method that relies on the registration of a healthy anatomical or probabilistic reference image to incorporate prior anatomical knowledge in the segmentation process and has been getting more and more attention in the recent years.

2.2 Brain Imaging

There exists a variety of different imaging modalities that enable the study of the brain. This section presents a brief overview of the different imaging methods, then focus on Magnetic Resonance Imaging (MRI) which is the most common modality for brain tumor observation.

2.2.1 Imaging modalities

Brain Imaging modalities can be grouped in two categories according to the information being captured by them: structural and functional imaging. The structural or anatomical modalities aim at visualizing the different structures and tissues of the brain. Among them, the most popular for Neuroimaging studies are the Computed Tomography (CT) and MRI. CT imaging relies on X-ray technology that is based on the absorption of X-rays beams as they pass through the different tissues of a patient's body. CT scans are constructed by using a series of X-ray beams that rotate around

the patient's head. Each beam yields a 2D image at a specific angle that are used to construct a 3D volume by tomographic reconstruction. CT scans are well contrasted and high resolution images. The main drawback of this technique is the the fact that it uses ionizing radiation.

MRI relies on the magnetic property of the hydrogen nuclei present in large quantity in the human body. It is the modality of choice for brain studies due to its high tissue contrast and details and non ionizing property. In the recent years, new techniques have developed such as Diffusion Tensor Imaging (DTI). This modality enables the reconstruction of the white matter tracts connecting the different parts of the neural networks in the brain by measuring the anisotropic diffusion of water inside the tissues. The presence of a tumor could have a direct impact on the fibers structures by causing disruption, displacement or infiltration of the fibers [Wei 2007]. Modalities such as MR spectroscopy (measurement of major metabolites in tumor tissue) and Perfusion MRI (measurement of the relative cerebral blood volume using a contrast agent) provide additional informations for the diagnosis and study of brain tumors [Soffietti 2010].

The goal of functional imaging is to study the human brain's function based on physiological changes caused by the brain's activity. Electroencephalography (EEG) and Magnetoencephalography (MEG) are techniques that offer means of directly measuring the brain activity. EEG detects the electrical impulses in the brain due to the neuronal activity via electrodes placed on the scalp. MEG measures the magnetic flux changes using sensors positioned closed to the scalp. The methods are popular due to their simplicity, the fact that they are non invasive and their very high temporal resolution. Determining precisely the spatial origin of the observed signal is however difficult.

Position Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) are nuclear imaging techniques that measure changes in the cerebral blood flow and tissue metabolism and how it is altered by brain disorders. A biological molecule is marked with a radioactive isotope and injected in the bloodstream and accumulated in areas where the molecule has affinity. The advantage of this method is that the radioactive tracer can be designed specifically to target specific organs or processes related to an illness. For instance, a radioactive isotope typically used is the Fluorodeoxyglucose (FDG) that behaves like glucose molecules and therefore is able to trace the brain's metabolic activity. However, the technique is invasive and potentially harmful due to the use of radioactive tracers while at the same time producing the isotopes is costly and difficult. Nowadays, CT and PET technologies are being combined in one device, allowing to combine the anatomical information recovered from the CT scans with the metabolic information given by the PET scans.

Functional MRI (fMRI) exploits the Blood Oxygen Level Dependent (BOLD) contrast [Ogawa 1990] to detect changes in the neuronal activity induced by sensorimotor or cognitive tasks. Neuronal activation causes an increase of the blood flow to compensate the oxygen consumption and therefore reduces the amount of deoxygenated hemoglobin molecules. The detection of changes in hemoglobin oxygenation relies on the paramagnetic properties of the deoxygenated hemoglobin which impacts the measured NMR signal. fMRI is often used for tumor surgery planning in order to identify the spatial relationship between the lesion and the functional area and evaluate the corresponding risks. Time constraints in the acquisition of the signal cause the fMRI images to be of

lower quality than the structural MRI.

Functional imaging is particularly important for brain tumor surgery planning. It enables to establish a relationship between the lesion and the functional area and plan the surgery accordingly (extent of resection, awake craniotomy...).

2.2.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging is one of the most popular medical imaging modalities. It has the advantage of being a non invasive and non ionizing technique that produces images with excellent tissue contrast, making it the modality of choice for brain (and more specifically brain tumors) study. MRI is based on the principle of Nuclear Magnetic Resonance (NMR), that is commonly used in spectroscopy to study the physical and chemical properties of molecules.

2.2.2.1 Physical principles

MRI relies on the magnetic properties (nuclear spin) of the hydrogen nuclei present in high quantities in the human body. The hydrogen nuclei spins around its axis, leading to the creation of an electromagnetic field. In natural condition, the spins' orientation is random and therefore, the total magnetic field is null. When placed into a large magnetic field B_0, the nucleis' spins align along (positive spin) or against (negative spin) the external field. The spins precess around the magnetic field's axis with a precession frequency or Lamor frequency ω_0 proportional to the external field:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

The magnetic vector of spinning nuclei can be expressed through the longitudinal component (M_z, parallel to B_0) and the transverse component (M_{xy}, perpendicular to B_0). In practice, the number of positive spins is slightly higher than the number of negative spins, yielding a longitudinal magnetization M_z. Since the spins do not precess in phase, the transverse magnetization is null.

When a radiofrequency (RF) pulse B_1 is introduced at the resonance frequency ω_0, there is an exchange of energy between the nuclei and the RF pulse which affects the spin equilibrium (phenomenon known as resonance). This causes the net magnetization vector to tilt away from the longitudinal axis and the creation of a transverse magnetization that can be measured by a receiver coil. In practice, the pulse B_1 is perpendicular to B_0, which causes the longitudinal component to be null at resonance and the transverse signal to be the strongest possible.

When the RF pulse is turned off, the nuclei progressively returns to equilibrium (relaxation phenomenon). During relaxation, the absorbed electromagnetic energy is retransmitted and constitute the NMR signal. Two different kinds of relaxation processes are observed: longitudinal and transverse. The longitudinal relaxation corresponds to the spin-lattice (surrounding tissue) interaction. The spin returns to its equilibrium state (alignment with the magnetic field B_0) by giving energy to the lattice. The recovery of the longitudinal magnetism follows an exponential curve and is characterized by a tissue specific time constant T1 after which 63 % of the final value is

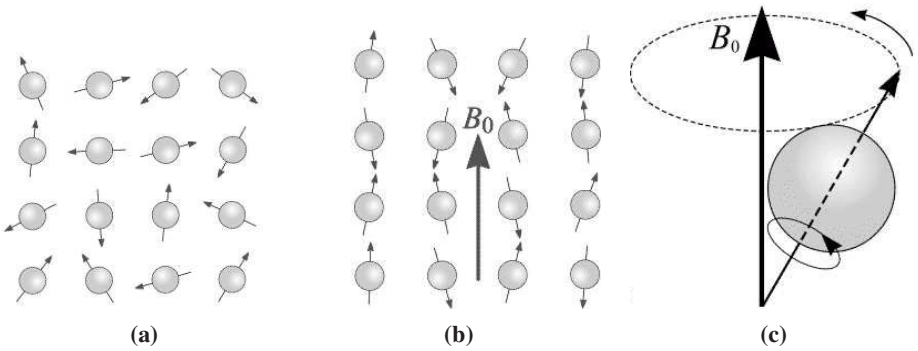


Figure 2.2: *Impact of an magnetic field B_0 on the spins' magnetization. (a) Random spins orientation, (b) alignment with the external field, (c) spin precession around B_0. Images taken from [Puddephat 2002]*

recovered.

$$M_z = M_0(1 - \exp(-\frac{t}{T_1})) \tag{2.2}$$

M_0 is the net magnetization at equilibrium that depends on the proton density and the strength of the external magnetic field. The transverse relaxation describes the spin-spin interaction and corresponds to the spins getting out of phase. This yields an exponential diminution of the transverse component characterized by a tissue specific time constant T2.

$$M_{xy} = M_{xy0} \exp(-\frac{t}{T_2}) \tag{2.3}$$

M_{xy0} is the amplitude of the transverse signal following the RF pulse. When the RF pulse is perpendicular to B_0, M_{xy0} is equal to M_0. In practice, T2 is always smaller than T1. The two constants are the key for the image contrast and definition of the different MRI sequences.

2.2.2.2 Image Sequences and DLGGs Imaging

An MRI sequence is characterized by two parameters. The Echo Time (TE) is the time between the RF pulse and the signal measurement, and the Repetition Time (TR) is the time between two RF pulses. After each RF pulse, the received NMR signal shows the evolution of the transverse component during relaxation. The signal oscillates at the resonance frequency and its envelope decays following an exponential rule. The initial amplitude corresponds to the magnitude of the longitudinal component and impacts its recovery status. The signal will be strong if the longitudinal component has been well recovered.

TR defines the time between two pulses and therefore the time allowed for longitudinal recovery. If TR is short, the longitudinal magnetization of tissues with a long T1 will not have time to recover completely and their initial signal will be weaker, while the signal of tissues with a small

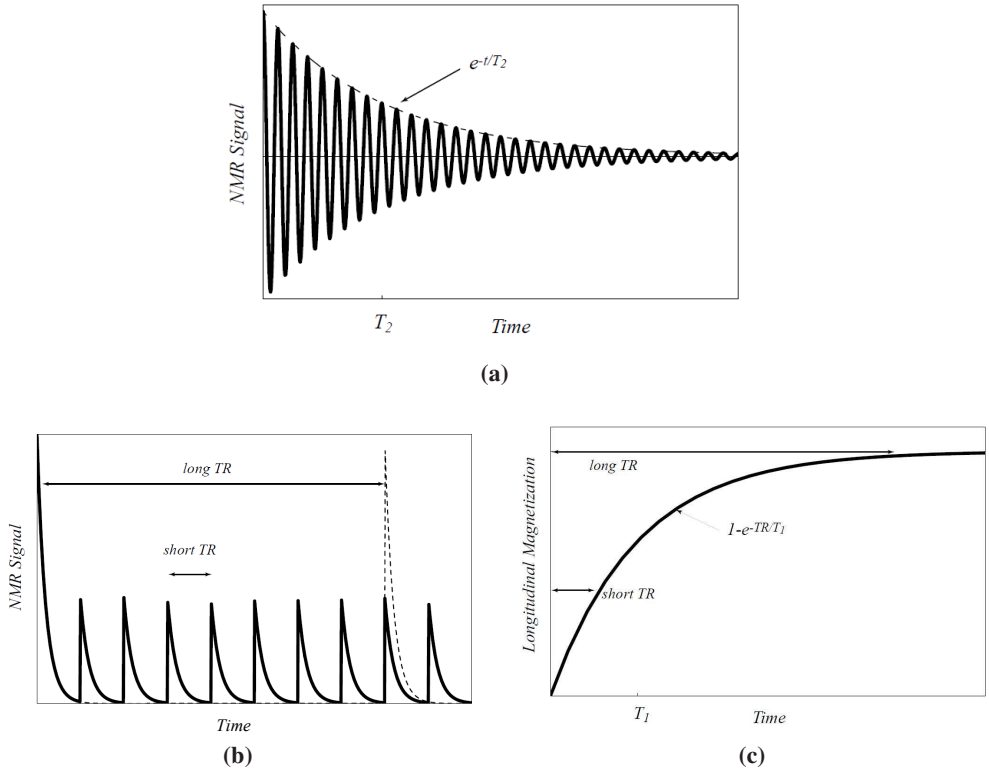
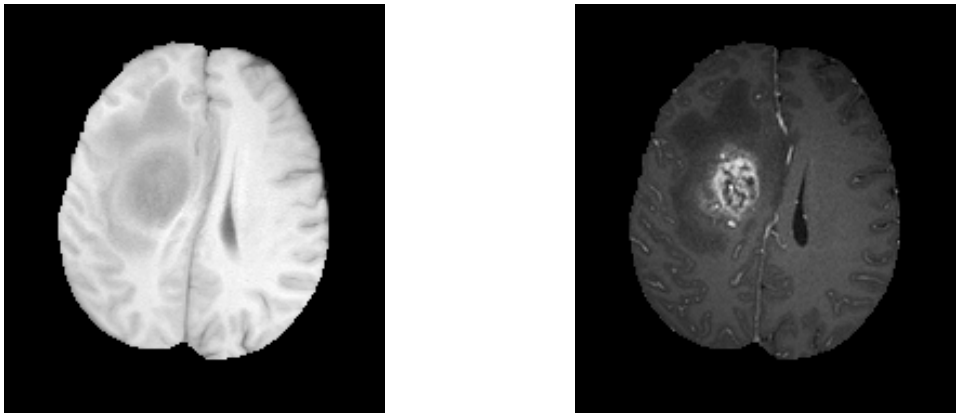


Figure 2.3: *Impact of TE and TR on the NMR signal [Buxton 2009]. (a) NMR signal, amplitude dependent on T2, (b) Impact of TR on the initial amplitude of the signal, (c) longitudinal magnetization dependent on TR and T1.*

T1 will be higher (large magnetization). This defines the T1-weighted modality, where the image contrast between tissues is defined by their T1 constant value. The image contrast no longer depends on TR when TR is set long enough so that complete longitudinal recovery is achieved for all tissues (same initial magnitude M_0). A long TE renders the image dependent on the transverse relaxation time T2. The signal will be weaker for tissues that have faster transversal decays (small T2 value). The T2-weighted images are obtained in this manner. Eventually, T1 and T2 are no longer discriminant parameters when TR is long and TE is short. This yields the Proton Density weighted image, where the strength of the signal depends on the strength of the acquired magnetization M_0 (dependent on proton density).

Brain tumor segmentation algorithms often exploit the different modalities to include maximal information. Visualization of brain tumor on T1 or T2 images alone is far from ideal. The tumors appear hypointense or normal on T1 images, and are therefore hard to detect and delineate. Better detection can be achieved by the injection of contrast agents, the most common being gadolinium contrast agents. It results in a tumor that is hyperintense in T1 images. However, complete en-



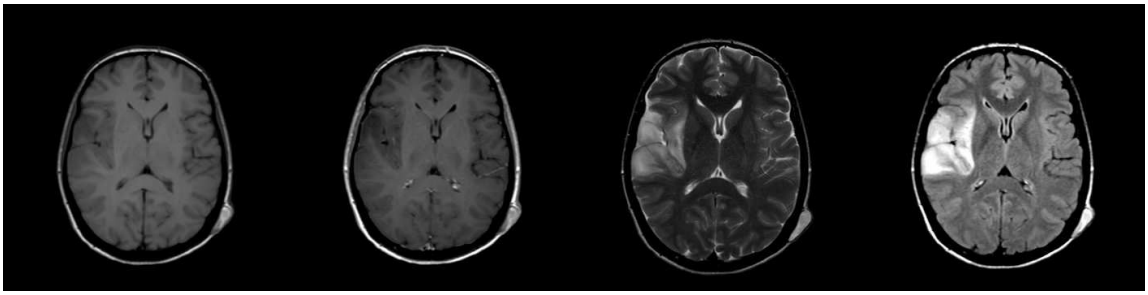
(a) **(b)**

Figure 2.4: *Contrast enhancement of High grade gliomas. (a) T1 weighted image, (b) contrast enhanced T1 weighted image.*

hancement is far from systematic especially in the context of DLGGs. About 20 % of DLGGs display a patchy and faint contrast enhancement. Contrast enhancement can also be associated to anapestic transformation. Therefore, methods that depend on contrast enhancement can only segment a subset of tumors and very few DLGGs. Contrarily to T1 images, the tumor appears hyperintense on T2 images and can be well visualized. The contrast is poor though, due to the high signal intensity of fluids which can be similar to the tumor's intensity.

Another popular sequence for the observation of brain tumors is the T2 FLAIR (Fluid Attenuated Inversion Recovery) sequence. This sequence is constructed by initially flipping the longitudinal magnetic field M_z in the opposite direction using an additional electromagnetic pulse oriented at 180° from the original field B_0. This pulse is followed by a longitudinal relaxation to return to the equilibrium, passing through the null value. An additional time constraint TI (Inversion Time) determines when the perpendicular RF pulse is added. It is possible to suppress certain tissues' signal by choosing TI so that it corresponds to the time when the signal magnitude is null. In FLAIR sequences, TI is set in order to suppress the high cerebrospinal fluid (CSF) signal in T2 images. The FLAIR sequence is one of the best contrasted modalities to visualize brain tumors and is the gold standard for DLGG observation [Bynevelt 2001].

An additional difficulty inherent to MR images is the problem of intensity inhomogeneity. The problem, caused by imperfections in the image acquisition process (instrumentation limitations or patient motion), is observed from one acquisition to the next and even within the same acquisition (bias field corruption)[Sled 1998]. This phenomenon can cause strong misclassifications when the segmentation technique relies on the observed intensity and/or on training data. Many algorithms have been developed in order to correct those inhomogeneities [Nyúl 1999, Vovk 2007], but the problem becomes even more difficult in the presence of a pathology since it can have similar visual



(a) (b) (c) (d)

Figure 2.5: *Images sequences for the visualization of a Low grade glioma. (a) T1, (b) T1 contrast enhanced (no enhancement of the tumor), (c) T2, (d) FLAIR images.*

behavior as bias field corruption.

2.3 Surface Based Segmentation: Deformable models

Deformable models have their origins on Snakes or Active Contours in [Kass 1988]. The object contour is represented as an explicit parametric surface $\mathcal{S} : \mathbb{R} \rightarrow \mathbb{R}^3$ that evolves from an initial configuration towards the optimal solution according to internal and external energy forces. The internal energy imposes smoothness and regularity of the curve while the external energy is based on the image features, guiding the curve towards locations of large gradient (object boundaries). The optimal contour is obtained by minimizing the energy functional. Despite the appeal and popularity of the concept, it suffers from several drawbacks, notably its sensitivity to initialization and tendency to get trapped in local minima, and inability to handle topological changes between the initial and the final configuration due to the explicit representation.

The level set method [Osher 1988] was introduced in medical imaging in [Malladi 1995]. It constitutes an elegant way of dealing with the inherent limitations of point-based snakes representations and in particular the handling of topological changes. It consists of an implicit representation of the object contour as the zero level set of a function $\Phi(\mathbf{x}, t)$ of higher dimension. The implicit function is designed so that it is positive outside the contour and negative inside, and commonly set as the signed distance from the contour.

$$\mathcal{S} = \{\mathbf{x} | \Phi(\mathbf{x}, t = 0) = 0\} \tag{2.4}$$

This yields an attractive formulation for shape modeling [Leventon 2000, Chen 2002, Rousson 2002] and image segmentation that handles topological changes as well as complex geometries and does not depend on the contour's parametrization. The evolution of the surface follows the surface's normal direction with a speed F, and is governed by a Partial Differential Equation:

$$\frac{\partial \Phi}{\partial t} = F(t) |\nabla \Phi| \tag{2.5}$$

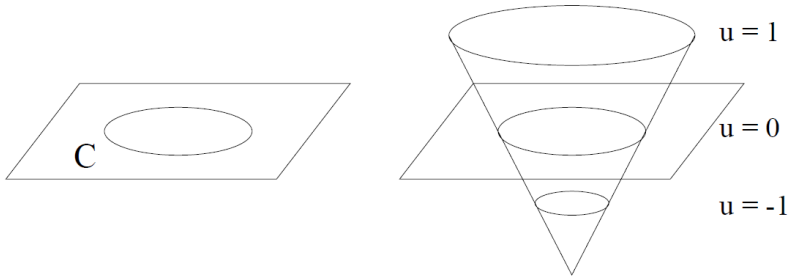


Figure 2.6: *Level set of an implicit function u and corresponding curve C [Leventon 2000].*

The design of the speed function F is an essential aspect of the framework and one of the strongest challenges of the method. Such a speed function can be either the natural outcome of a minimization process or can be directly introduced depending on the nature of problem being considered. In the most general case it is the sum of two terms: an image term F_{Im} and a curvature term.

$$F(t) = F_{im} + \gamma \nabla \frac{\nabla \Phi}{|\nabla \Phi|} \tag{2.6}$$

The mean curvature term is a smoothness term which aim is to prevent leakage of the level set in noisy and indistinct areas. The parameter γ determines the amount of smoothing and is of key importance as a value too high would lead to under segmentation.

Despite extensive work in the medical imaging domain, the use of level sets for the segmentation of brain tumors is limited. The main drawbacks of the level set framework is its sensitivity to initialization and parameter setting as well as the computational burden of solving the partial derivative equation. In order to enable interactive parameter tuning, Droske et al. [Droske 2001] proposed an adaptive multi grid framework allowing for almost real time feedback on the curvature terms and front evolution speed. They adopt the Fast Marching Algorithm [Sethian 1996]. Assuming the speed function is always positive and the surface is initialized inside the tumor, the evolution of the contour will always be outwards. This allows to compute the arrival time $T(\mathbf{x})$ of the front on a voxel $\mathbf{x}$. This converts the problem into solving the so-called Eikonal equation:

$$|\nabla T| F = 1 \tag{2.7}$$

The method is accelerated by considering an adaptively sampled grid (the highest resolution corresponding to the voxel level) based on the local image homogeneity evaluated using the image gradient. Similarly, Lefohn et al. [Lefohn 2003] used GPU programming to reduce the running time, enabling interactive handling of the free parameters. Both methods used simple image terms based on image gradient and contrast that are sensitive to image noise and unclear boundaries, which enforces the need for expert supervision.

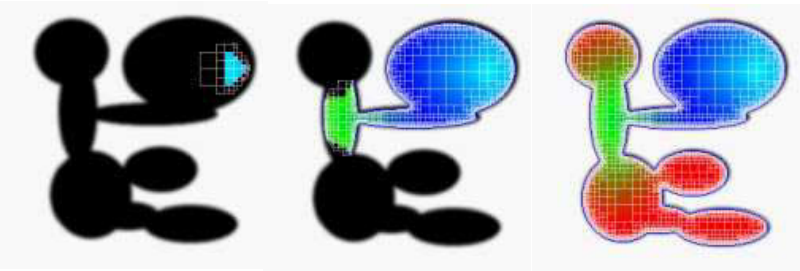


Figure 2.7: *Figure extracted from [Droske 2001]. Adaptive grid sampling, progressively constructed as the front progresses. The colors indicates the arrival time of the propatation (from blue (early) to red)*

Ho et al. [Ho 2002] proposed an fully automatic 3D level sets framework that relies on region based methods [Zhu 1996, Chan 2001, Paragios 2002]to avoid manual initialization and parameter tuning. An initial separation of the (enhanced) tumor and background pixels is performed by fuzzy classification, used for initialization of the surface. Classification is performed by fitting two distributions to the histogram of the pre and post gadolinium enhanced T1 weighted MRI images difference.The derived probabilities are then used to guide the evolution of the level set function in a region competition manner:

$$F_{im} = \alpha(P_{tum}(\mathbf{x}) - P_{bg}(\mathbf{x})) \tag{2.8}$$

The method was applied to glioblastomas and meningiomas. The performance of the method depends heavily on the quality of the classification and while it also requires contrast enhancement.

A similar region competition level set framework was presented in [Popuri 2012]. A series of features are extracted from the images (multi modal intensities, Gabor features for texture information and brain hemispheres symmetry) and used to perform an initial classification of the image using K-means clustering [MacQueen 1967] (see section 3.2.2). The contour evolution is guided by the clusters value probabilities combined with a Dirichlet prior which role is to penalize clusters that are predominant in the healthy brain and wrongly assigned to the tumor region. A post processing step to remove CSF and skull misclassification was added. The method was tested on 15 multi modal MRIs (T1, T1 contrast enhanced (T1CE), and T2).

Taheri et al. [Taheri 2010] combined adaptive image thresholding with level sets segmentation. The level set was initialized inside the tumor area, leading to an initial guess of the threshold value. The threshold value was then iteratively updated based on the current segmentation. The contour evolves based on the signed intensity difference with respect to the threshold value, and stops when the variation of the threshold from one iteration to the next is lower than a predefined parameter. This approach assumes relatively strong intensity difference between tumor and background.

Deformable models are attractive methods that have proven to handle complex geometries and produce smooth and precise segmentations. They suffer however from several drawbacks that reduce their interest for the difficult problem of tumor segmentation. Their main weakness is their

sensitivity to initialization and parameter selection. Interactive parameter tuning and initialization is promising, but implies extensive user work and is dependent on the user's knowledge. The combination with region based approaches enables automation but renders the final segmentation strongly dependent on the quality of the classification scores.

2.4 Region-Based Segmentation

Region-based segmentation methods aim at directly identifying the tumor voxels instead of finding the tumor's surface. We distinguish unsupervised classification that is solely based on the image's appearance and properties, from supervised classification where the separation criteria are learned from a labeled training set. Both approaches have been considered for tumor segmentation and will be reviewed in this section.

2.4.1 Classification methods

2.4.1.1 Unsupervised Classification and Knowledge based Methods

Unsupervised classification relies on intrinsic properties of the samples. Those methods were among the first considered due to their simplicity and the fact that classification is performed solely based on the observed samples without the need for labeled training data. The concept of clustering is to regroup observations into groups (clusters) with similar properties. Early work in tumor segmentation relied on clustering to classify the image into different categories. We distinguish hard clustering where each sample belongs to only one cluster from fuzzy clustering, where each sample is assigned a membership score for each cluster.

Phillips et al. [Phillips 1995] proposed fuzzy clustering to perform automated glioblastoma segmentation. The method adopted is the Fuzzy-C-means (FCM) algorithm [Dunn 1973, Bezdek 1981]. This is an unsupervised clustering technique derived from the well known hard clustering K-means algorithm, with the difference that the cluster membership is probabilistic. Fuzzy partitioning is carried out by minimizing of the sum of weighted distances between the samples and the cluster centers:

$$J_m = \sum_{1}^{n}\sum_{1}^{K} u_{ij}^m \|\mathbf{x_i} - \mathbf{c_j}\|^2 \tag{2.9}$$

where K is the desired number of clusters (tissue classes), $\mathbf{c_j}$ is the center of cluster j and u_{ij} is the degree of membership of sample $\mathbf{x_i}$ in cluster j that verifies:

$$\begin{aligned} \sum_{i=1}^{C} u_{ij}^m &= 1 \ \forall j \\ \sum_{j=1}^{n} u_{ij}^m &> 0 \ \forall i \end{aligned} \tag{2.10}$$

After initialization of the memberships, the clustering is obtained via an iterative process, alternating between computation of the cluster centers (as the mean of all points weighted by their membership) and of the cluster memberships. The samples were clustered based on their multispectral intensities values (T1, T2 and Proton Density). Numerous misclassifications were observed, which revealed the strong intensity overlap between the healthy tissue and the tumor as well as the insufficiency of the algorithm alone to perform meaningful segmentation.

Clark et al. [Clark 1998] proposed a knowledge based automated segmentation method on multispectral MRI data for the 2D segmentation of enhanced glioblastomas. After learning an ensemble of rules on a training set of 17 slices, an initial segmentation of each slice is performed using Fuzzy C-Means clustering. The predefined rules enable to identify the abnormal slices to be further processed. After adaptive histogram thresholding based on the pixels mutli spectral intensities, the obtained segmentations are post processed based on the different heuristics. A similar system was developed in [Fletcher-Heath 2001] for non enhancing tumors by defining a different set of heuristics.

The quality of the segmentation depends strongly on the definition of the knowledge system, which should be general yet precise enough to obtain good segmentations for a broad range of tumors. The need for important expert guidance to define the system as well as the limited success on noisy or heterogeneous tumor have reduced the appeal of such approaches and highlighted the need for more elaborate methods.

2.4.1.2 Supervised Classification

Despite the theoretical and practical value of unsupervised classification, it has quickly shown limitations in the separation of object with complicated and heterogeneous appearances such as tumors. As a result, supervised classification techniques have become increasingly popular in the segmentation field. The idea is to learn a classifier from a set of pre labeled data and an ensemble of image features. Consider an ensemble of training pairs $\{\mathbf{x_i}, y_i\}$ where $\mathbf{x_i}$ corresponds to an observation (typically an image voxel) and y_i to a label ($\{-1, 1\}$ in the context of binary classification). Each pair is associated to a feature vector $f(\mathbf{x_i})$. Based on these training samples, a mapping function is learned and constitutes what is called a classifier. In its simplest form, the set of image features can the image intensities. Supervised classification involves a training phase during which the classifier is learned based on the labeled training set, followed by a testing phase during which a new unlabeled sample is assigned a label based on its features and the classifier's rules.

Vinitski et al. [Vinitski 1997] used multispectral (T1,T2 and Proton density) k Nearest Neighbor (kNN) classification for the segmentation of benign and malignant brain tumors. Each voxel is associated with a 3 dimensional vector corresponding to the intensities observed in the 3 different modalities. The training set consists of samples from the scan to be segmented (voxels) that are manually selected and labeled by an operator. A probability per class is then assigned to each unlabeled voxel based on the classes of its k closest samples, the k parameter being set manually. This relatively simple method requires manual selection of samples on every image to segment and

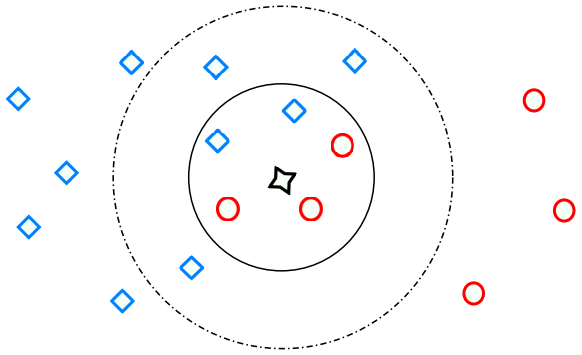


Figure 2.8: *Illustration of the kNN clustering approach. We aim at classifying the black sample as either a blue square or red circle. Setting k=5 selects the circle class, while k=8 (dashed circle) selects the other class.*

assumes homogeneity of the classes. Furthermore, it is sensitive to the choice of k, as illustrated in Fig. [2.8].

The most popular supervised classification method for brain tumor segmentation has undoubtedly been the Support Vector Machines (SVM) [García 2004, Verma 2008, Zhang 2004]. Introduced in [Boser 1992] for linear binary classification, it has become main stream in machine learning and pattern recognition. Given a set of training samples $\{\mathbf{x_i}, y_i\}$, the SVM method seeks a hyperplane $\{\mathbf{x}| < \mathbf{w}, \mathbf{x} > +b = 0\}$ defined by its normal vector $\mathbf{w}$ and the offset b, that separates the two different classes. The optimal hyperplane is chosen so that it maximizes the margin between the plane and the closest samples. The Support Vectors are the samples that are the closest to the hyperplane and therefore determine its position. In practice, the data is rarely linearly separable. Non linear separation is performed by mapping the data into a higher dimensional feature space $\mathbf{x} \rightarrow \Phi(\mathbf{x})$. The optimal hyperplane is then found by minimizing the following objective function:

$$\begin{aligned} &\min_{\mathbf{w},\xi,b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{1}^{n}\xi_i \\ &s.t.\ y_i(\mathbf{w}\cdot\Phi(\mathbf{x_i}) - b) \geq 1 - \xi_i \end{aligned} \tag{2.11}$$

where ξ_i are slack variables introduced to allow mislabeled samples, perfect separation being rarely possible when dealing with noisy data such as medical images. This optimization problem is solved using the Lagrangian formulation, where unknown variables α_i (lagrange multipliers) are introduced for each constraint. The support vectors are the samples that have non zero weights α_i.

$$L_n = \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \Phi(\mathbf{x_i})\Phi(\mathbf{x_j}) \tag{2.12}$$

Finding the appropriate mapping can be difficult, while the complexity of the algorithm augments with the size of the feature space. It becomes unnecessary when defining a kernel function

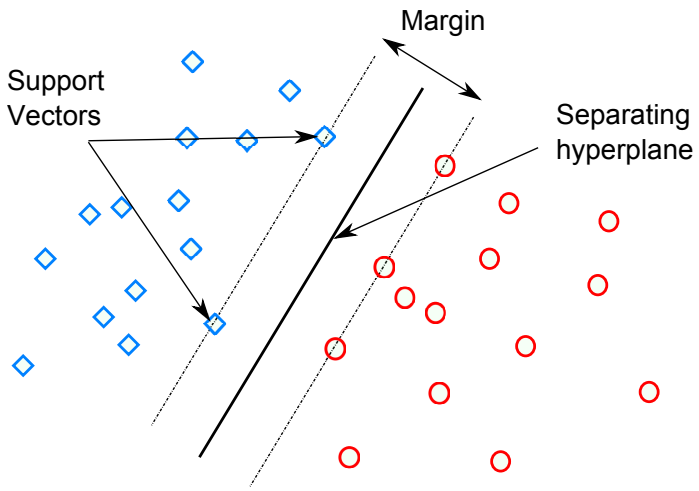


Figure 2.9: *Illustration of the SVM approach.*

$K(\mathbf{x_i}, \mathbf{x_j}) = \Phi(\mathbf{x_i})\Phi(\mathbf{x_j})$ that directly computes the dot product of the feature vectors. Using such a formulation, the problem becomes solely dependent on the number of samples, it is therefore possible to operate with a feature space of infinite dimension. Eventually, the classification decision of a new sample is recovered as:

$$h(\mathbf{x}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{2.13}$$

The SVMs being by definition binary classifiers, extension to multiclass classification is not straightforward. Current methods either use a hierarchical approach or a one vs all approach [Duan 2005].

Another popular classification algorithm is the AdaBoost algorithm. Boosting is a general machine learning method which goal is to improve the accuracy of any learning algorithm. AdaBoost (short for Adaptive Boosting) is the most popular boosting algorithm and was introduced in [Freund 1997]. It relies on the concept that a classifier with high discriminative performance (strong classifier) can be constructed as a weighted linear combination of weak classifiers. The most commonly used weak classifiers' structures are decision stumps or tree structured classifiers. The simplest classifier is preferred to maintain the speed of the algorithm. The weak prediction rules are learned iteratively by minimizing the classification error with an increased focus on misclassified samples at each new iteration. Several variants of the Adaboost algorithm have been developed, among them the Gentle Adaboost algorithm [Friedman 2000] that will be further detailed in section 3.3.1. The popularity of the algorithm can be explained by its simplicity, flexibility and speed. An example of AdaBoost being applied for tumor segmentation can be found in [Xuan 2007].

Recent work on Decision Forests (or Randomized Forests) showed strong promises [Zikic 2012, Geremia 2012]. The concept was introduced in [Breiman 2001] and corresponds to a collection of

tree structured classifiers (decision trees). A decision tree consists of a set of binary rules that progressively split the data into smaller sets based on their features values. Starting at the root node, the complete training data X is split and assigned to either the left X_L or right X_R child node. The split is chosen among a randomly selected set of possible splits regarding the feature values so that it maximizes the information gain:

$$I = -\frac{X_R}{X}H(X_R) - \frac{X_L}{X}H(X_L) \tag{2.14}$$

where $H(.)$ is the entropy calculated from a histogram of class values. The same process is repeated until the tree reaches a predefined depth d. Each leaf node (terminal node of the tree) is associated with a probability that corresponds to the fraction of the data of class c that has been assigned to the node. Several decision trees are learned in a similar fashion and the forest consists of the ensemble of those trees. During testing, the test data is pushed through all the decision trees by applying the different separation rules. Each tree yields a probability per sample and class based on the reached leaf node. The forest probability in then computed as the average probability over all the trees. Decision forests are inherently multi-class classifiers, robust to overfitting and fast, making them an appreciated classification approach.

A key element of the SVM, Adaboost or Decision forest algorithms is the design of the feature vector that should discriminate between healthy and diseased samples. The most straightforward features are intensity based, such as the multi modal intensities, intensity patches around the sample (to incorporate neighborhood information) on which first order statistical moments are often computed (mean, variance or entropy to name a few). Features related to the image texture have also been popular, among them the Haralick features [Haralick 1973], or the Gabor features [Manjunath 1996]. Xuan et al. [Xuan 2007] exploit the symmetry between the brain hemispheres by computing the intensity difference between the left and right hemisphere. Zikic et al. [Zikic 2012] use preliminary classification results obtained by modeling the data as a mixture of Gaussians and context aware features such as the intensity range along a 3D line. In the specific case of SVMs, kernel mapping can be used instead of (or combined with) the feature vector.

Despite the promises of supervised pattern classification methods, they are constrained by the i.i.d (independent and identically distributed) assumption that treats all samples independently. This leads to noisy and irregular segmentations that could be barely compensated by the introduction of neighborhood or context aware information in the feature vector or post processing with morphological operators. Those limitations have motivated the use of more elaborated models that take into account the spatial interaction between neighboring pixels. Graph based approaches offer an elegant and flexible way of doing so, modeling the spatial interactions as edges connecting the pixels.

2.4.2 Graph based Segmentation

The main idea behind graph based segmentation is to consider the image as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, that consists of a set of nodes $\mathcal{V} = \{1, 2, .., N\}$, corresponding to each voxel of the

image, and a set of edges $\mathcal{E} = (i, j) \in \mathcal{V}$, that represent the dependencies between the nodes. Typically, the edges only connect the adjacent nodes (4-neighborhood in 2D, and 6-neighborhood in 3D).

2.4.2.1 Affinity based Segmentation

The most natural way of exploiting such a graphical formulation is to partition the graph into connected components of similar properties. In such methods, the edges encode an affinity measure $w_{i,j}$ that represent the similarity between the connected nodes i and j. The problem of graph partitioning can be seen as the search for m disjoint sub graphs $\{\mathcal{V}_1, .., \mathcal{V}_m\}$, m being the number of sought classes, that verify $\mathcal{V} = \cup_{i=1}^{m} \mathcal{V}_i$ and $\cap_{i=1}^{m} \mathcal{V}_i = \emptyset$. The subgraphs are constructed by cutting the edges between them and constitute ensembles of vertices with strong affinities.

Considering the binary case, the cost of partitioning the graph into subsets A and B can be computed as the total affinity of the removed edges, called the cut of the graph:

$$cut(A, B) = \sum_{i \in A, j \in B} w_{i,j} \tag{2.15}$$

The optimal bipartition of the graph minimizes this cut value. The minimum cut method was used in [Wu 1993] for image segmentation as a clustering problem, seeking the smallest $k - 1$ cuts among all possible cuts to form k disjoint subgraphs. Despite promising results, it highlighted the tendency of the method to favor smaller subgraphs leading to inadequate segmentations. This led to the introduction of Normalized Cuts [Shi 2000] that compute the cost of a cut as a fraction of the total edges connections to all nodes in the graph:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, \mathcal{V})} + \frac{cut(A, B)}{assoc(B, \mathcal{V})} \tag{2.16}$$

where $assoc(A, V) = \sum_{i \in A, j \in \mathcal{V}} w_{i,j}$ is the sum of the connection weights from all nodes in A to all the nodes in the graph. The optimal Ncut can be approximated by eigenspectrum analysis [Shi 2000].

Affinity based segmentation has drawn limited interest for tumor segmentation. [Moonis 2002, Liu 2005] proposed interactive affinity based segmentation methods that rely on fuzzy theory. After manual selection and labeling of seeds, the affinity between neighboring voxels is computed. The smallest affinity value represents the strength of connectivity of a path from voxel i to voxel j. Eventually, the strength of connectivity between both voxels is set as the largest path among all possible paths from i to j. The tumor is segmented as the voxels that are connected to the tumor seeds with a strength larger than an empirically chosen threshold.

Corso et al. [Corso 2008] proposed a hybrid approach between Bayesian classification and graph partitioning for segmentation of glioblastomas. The affinities are defined in a probabilistic fashion so that they are class dependent:

$$P(X_{ij}|i, j) = \sum_{c_i} \sum_{c_j} P(X_{ij}|i, j, c_i, c_j) P(i|c_i) P(j|c_j) P(i, j) \tag{2.17}$$

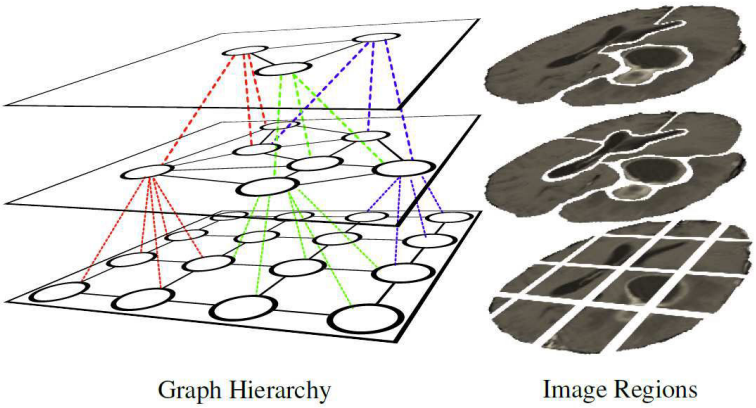


Figure 2.10: *Illustration of the SWA algorithm extracted from [Corso 2008].*

where X_{ij} is the binary event describing if nodes i and j are assigned to the same class. The likelihood probabilities $P(i|c_i)$ are learned as mixtures of Gaussians and lower the affinity if the class assignment is unlikely. The first term $P(X_{ij}|i, j, c_i, c_j)$ is the affinity function weighted by a class dependent parameter learned over a training set. Graph partitioning relies on the Segmentation by Weighted Aggregation (SWA) algorithm, a multilevel hierarchical approach proposed in [Sharon 2006] for Normalized cuts partitioning. An initial high resolution graph $\mathcal{G}^0$ (one node per voxel) is progressively coarsened by aggregating nodes that have strong affinities. Weights are updated by interpolation and edges are constructed if the weights are positive. The algorithm stops when the coarsened graph is made of one single node or if there are no connections between the remaining nodes. At each level, structures of different scales are extracted. In [Sharon 2006] they are recovered by associating each voxel with the most salient segment (lowest normalized cut) of which it is a part in the pyramid. This segment recovery method has been that successful in the context of medical images. Corso et al. [Corso 2008] proposed computing the maximum likelihood at each level weighted by the inter-level interpolation weights. The method was tested for the segmentation of multi spectral glioblastoma images (T1 pre and post contrast, T2 and FLAIR) into 4 classes (tumor, edema, brain and background).

2.4.2.2 Markov Random Fields

MRFs where first introduced into vision in [Geman 1984] for image restoration, and have since been one of the most popular undirected graphical model in the domain. They have been successfully applied to a broad range of applications [Kolmogorov 2002, Boykov 2006, Komodakis 2007b, Glocker 2008a].

Let $\mathbf{X} = \{X_1, ..., X_n\}$ be a set of random variables defined on $\mathcal{V}$, each variable X_i taking a value l_i in a label space $\mathcal{L}$. A joint event $\{X_1 = l_1, ..., X_n = l_n\}$ is called a configuration and

abbreviated as $\mathbf{X} = \mathbf{l}$, where $\mathbf{l} = \{l_1, ..., l_n\}$. We denote $P(X = \mathbf{l})$, , abbreviated as $P(\mathbf{l})$, the joint probability of a configuration. A random field is a Markov Random Field on a graph $\mathcal{G}$ if the two following conditions are satisfied:

$$\begin{aligned} \text{Positivity: } & P(\mathbf{l}) \geq 0 \; \forall l \in \mathcal{L} \\ \text{Markov Property: } & P(l_i | l_{\mathcal{V}-\{i\}}) = P(l_i | l_{\mathcal{N}(i)}) \end{aligned} \tag{2.18}$$

$\mathcal{N}(i)$ is the neighborhood system defined as: $\mathcal{N}(i) = \{j | i, j \in \mathcal{E}\}$. The Markov property imposes that the conditional distribution of a variable is only dependent on its neighbors. The problem of image segmentation is reformulated as a labeling problem, where the goal is to assign a label to each node (voxel). The labels correspond to the different image classes. For binary classification, the label set is defined as $\mathcal{L} = \{0, 1\}$. The optimal labeling should maximize the joint configuration probability:

$$\hat{\mathbf{l}} = \underset{\mathbf{l}}{\operatorname{argmax}} \, P(\mathbf{l}) \tag{2.19}$$

The Hammersley-Clifford theorem [Besag 1974] establishes the equivalence between an MRF and a Gibbs distribution, allowing to link the local conditional probability to a global probability on the whole distribution:

$$P(\mathbf{l}) = \frac{1}{Z} \exp\left(-\frac{1}{T} \sum_{C} V_C(l_C) \right) \tag{2.20}$$

Where T is a constant called the temperature (holdover from the formulation's origins in statistical physics) typically set to 1 and Z is a normalizing constant called the partition function:

$$Z = \sum_{\mathcal{L}} \exp\left(-\frac{1}{T} \sum_{C} V_C(l_C) \right) \tag{2.21}$$

C denotes a clique that is a fully connected subgraph and $V_C(.)$ is an energy function defined on the subset defined by the clique, called the clique potential. The estimation of the partition function is computationally intractable. It is not critical for inference, but becomes important for learning since it depends on the model's parameters. It can be avoided for inference by working in the energy domain and minimizing the energy function:

$$E(\mathbf{l}) = \sum_{C} V_C(l_C) \tag{2.22}$$

Considering cliques of size up to two leads to the well known pairwise MRF energy:

$$E(\mathbf{l}) = \sum_{i \in \mathcal{V}} V_i(l_i) + \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}(i)} V_{i,j}(l_i, l_j) \tag{2.23}$$

MAP-MRF Framework

The MAP-MRF framework relates to Bayesian statistics, where one seeks an optimal labeling given an observed image $\mathbf{I}$ by maximizing the Maximum a Posteriori probability $P(\mathbf{l}|\mathbf{I})$. Following Bayes rule, the posterior probability is linked to the joint probability as:

$$\hat{\mathbf{l}} = \arg\max_{\mathbf{l}} P(\mathbf{l}|\mathbf{I}) \propto P(\mathbf{I},\mathbf{l}) = P(\mathbf{I}|\mathbf{l})P(\mathbf{l}) \tag{2.24}$$

where the prior probability $P(\mathbf{l})$ is modeled as an MRF. Typically, only pairwise cliques are considered, but higher order cliques can also be considered to incorporate more elaborate prior knowledge. Conditional independence of the voxels is assumed given their label class ($P(\mathbf{I}|\mathbf{l}) = \prod_i (P(I(i)|l_i))$ [Besag 1986]. The MAP estimate of the posterior probability is equivalent to minimizing the corresponding energy function:

$$E(\mathbf{l}) = \sum_i -log(P(I(i)|l_i)) + \sum_i \sum_{j \in \mathcal{N}(i)} V_{i,j}(l_i, l_j) \tag{2.25}$$

The singleton potential is the data term or likelihood potential that measures the cost of assigning a label. The pairwise potential is a regularization prior ensuring local consistency and smoothness of the labeling. The MRF formulation allows flexibility in the design of the energy potentials, which are the key aspects of the model. One of the most simple and popular spatial interaction model is the homogeneous and isotropic Potts model (or Ising model for binary labelings) that penalizes different segmentation labels between neighbors:

$$V_{i,j}(l_i, l_j) = \beta(1 - \delta(l_i, l_j)) \tag{2.26}$$

where $\delta(.,.)$ is the Kronecker delta function and the parameter β determines the amount of regularization. The likelihood potential is usually modeled using pattern classification techniques, the role of the MRF prior being to counterpart the i.i.d assumption of those approaches. Conditional Random Fields (CRF) [Lafferty 2001] differ from MRFs by the fact that they are globally conditioned on the observation $\mathbf{I}$. They are thus discriminative models that directly model the conditional posterior probability, allowing to incorporate information from the observed variables with more flexibility. Notably, it allows the definition of a pairwise regularization term that is dependent on the observation.

MRF Optimization

Inference of MRFs/CRFs is a challenging task and many algorithms have been developed to tackle the problem. Among the earliest methods, we can cite the Iterated Conditional Modes (ICM) [Besag 1986] and simulated annealing (SA) [Kirkpatrick 1983]. ICM performs iterative local optimization of the posterior probability. This makes a relatively efficient algorithm that is however strongly sensitive to initialization. Simulated annealing methods are, on the contrary, more likely to reach global optima but computationally intractable methods. Nowadays, efficient methods have been developed that can be separated in two categories: graph cuts based approaches and message passing methods.

Graph cuts convert the problem of pairwise MRF optimization for binary labeling into a graph partitioning one by the construction of a directed graph (the s-t graph) where two terminal nodes (source s and sink t) corresponding to the two different labels have been introduced. An s-t cut separates the graph into two disjoint sets S and T, each connected to one terminal node. Nodes are assigned the label associated to the terminal node they are connected to. The graph is constructed so that the cost of an s-t cut (sum of removed edges' weights) is equal to the MRF energy given the corresponding configuration l. The problem is rendered equivalent to the min-cut problem introduced in section 2.4.2.1. The theorem of Ford and Fulkerson [Ford 1962] is an important result that states that finding the min-cut is equivalent to finding the maximum flow from the source s to the sink t (max-flow problem). The max-flow saturates a set of edges that corresponds to the min-cut. Low complexity methods have been proposed to solve the min-cut/max-flow problem, among which augmenting path methods [Ford 1962] and push-relabel methods [Goldberg 1988]. A review and comparison of both approaches can be found in [Boykov 2004]. Two limitations of the method are the restriction to binary labels and the requirement that the pairwise potentials are submodular in order to reach a global optimum:

$$V_{pq}(0,0) + V_{pq}(1,1) \leq V_{pq}(0,1) + V_{pq}(1,0) \tag{2.27}$$

For non submodular functions, a method referred to as Quadratic Pseudo-Boolean Optimization (QPBO) was considered [Kolmogorov 2007]. QPBO constructs an s-t graph where each node in the original MRF is represented by two complementary nodes $\mathbf{x_i}$ and $\bar{\mathbf{x}}_\mathbf{i} = 1 - \mathbf{x_i}$. Solving the graph's min-cut/max-flow problem yields three types of label: 0,1 or unknown if the label of nodes $\mathbf{x_i}$ and $\bar{\mathbf{x}}_\mathbf{i}$ is inconsistent. QPBO provides a partial solution with guarantee that the labeled variables correspond to the optimal solution.

Extensions to the multi label case were proposed in [Boykov 2001]. They proposed the α-expansion and $\alpha - \beta$ swap methods that iteratively optimize the MRF energy by defining a set of possible moves based on the initial MRF configuration at each iteration, and select the configuration that locally minimizes the energy. α-expansion allows any node to change its label to α and only α. $\alpha - \beta$ swap only allows nodes labeled α to change their label to β and vice versa. Each iteration corresponds to a binary s-t cut computation. The local minimum recovered by α-expansion is guaranteed to be within a known factor of the global optimum.

In [Komodakis 2007b, Komodakis 2008b] a linear programming based method called Fast-PD was proposed for optimization of MRFs. It relies on linear programming relaxation and primal-dual gap minimization. The method can be seen as a generalization of the α-expansion method and is substantially faster. The algorithm is used for MRF optimization throughout this thesis and a short review of the method will be provided in the next chapter.

Message passing techniques propagate messages across the nodes to evaluate the optimal configuration. Belief Propagation (BP) [Pearl 1988] is an iterative algorithm where at each iteration, a node $\mathbf{p}$ sends a message $m_{pq}(l)$ to node $\mathbf{q} \in \mathcal{N}(\mathbf{p})$ describing how likely the assignment of label l to node $\mathbf{q}$ is. The message is computed based on the messages node $\mathbf{p}$ previously received and the local energy costs. Two kinds of BP algorithms exist: the max-product that estimates the MAP configuration and the sum-product that computes the marginal probabilities. BP is exact for tree structured graph where messages are passed from the leafs to the roofs. It is approximate where the graph contains loops (loopy belief propagation) and there is no guarantee of convergence.

Sequential Tree Reweighted message passing (TRW-S) [Kolmogorov 2006] is a message passing algorithm strongly related to linear programming relaxation. Here, the solution is sought by maximizing the dual problem (lower bound) of the relaxed linear integer program. This is done by splitting the problem into convex subproblems defined as trees. Then, the algorithm considers each node sequentially and performs belief propagation in each tree it belongs to followed by an averaging step. In [Komodakis 2007a] a new message passing method was proposed based on dual decomposition. It relies on decoupling the Lagrangian dual problem into several subproblems through Lagrangian relaxation. The decoupled subproblems (the "slaves", defined as subtrees of the MRF graph) are easily solved using the max-product BP algorithm. The method iterates between solving the simple subproblems and sending the obtained minimizers to the "master" which role is to update the slaves parameters (using the projected subgradient method) so that the dual is maximized.

Applications to Tumor Segmentation

An interactive hierarchical MRF model was presented in [Gering 2002]. An initial intensity based segmentation at the voxel level is progressively refined by incorporating higher level knowledge. The voxel based segmentation relies on estimation of the MRI bias field using the Expectation-Maximization (EM) algorithm [Wells III 1996b]. The EM algorithm [Dempster 1977] is an iterative procedure for the estimation of the maximum likelihood of probabilistic models in the presence of missing data (also called latent variables). In this work, the missing data are the tissue classes. The healthy classes intensities (likelihood) are modeled as Gaussian distributions, while the tumor class is modeled as a deviation from normalcy (minimum intensity distance from all other classes). The bias field is modeled as a zero-mean additive Gaussian distribution. The EM algorithm iteratively alternates between two steps: the E-step corresponds to the estimation of the posterior probabilities given the currently estimated parameters (bias field and Gaussian parameters here) while the M-step is a MAP estimation of the bias field given the estimated tissue probabilities. Initialization of the Gaussian parameters is done by the user by selecting representative voxels of

each class.

The next level refines the obtained classification using an MRF model, whose parameters are learned from training data. Additional multi level MRF models are then used to refine the classification results based on structure size (expected distance from the structures boundaries) and inter structure distances.

In [Wels 2008], the likelihood term was constructed by Probabilistic Boosting Trees classification (PBT) [Tu 2005] and smoothed with an Ising prior weighted by the relative voxels' intensities. The idea of PBT is to progressively diminish the complexity of the learning process through a recursive construction of the classification tree. At each tree node, a strong classifier is learned using the AdaBoost algorithm. The data is then split and assigned to two child nodes based on their classification results. This progressively reduces the heterogeneity of the data at each node and simplifies the classification task. The method was optimized using Graph Cuts and tested for 3D segmentation of 6 multi spectral (T1, T1CE, T2) pediatric MR images.

Lee et al. [Lee 2005] proposed to combine SVM classification with CRF smoothing for tumor segmentation. The smoothing term was modeled as a penalizing constant between neighboring nodes weighted by the nodes' feature vector computed for the SVMs:

$$V_{i,j}(l_i, l_j, \mathbf{x})) = l_i l_j \nu^T \gamma_{i,j}(\mathbf{x}) \tag{2.28}$$

where $\gamma_{i,j}(.)$ is the weighting function describing the similarity between nodes i and j, and ν^T is a node dependent parameter that has to be inferred. The idea is to encourage discontinuities between neighboring nodes if they display different properties. After learning of the CRF parameters, the optimization was performed using the ICM approach. The method was tested on 7 patients using patient specific training for the classifier (the training set consists of extracted slices of the image to be segmented). To cope with the difficult and time consuming task of learning a CRF's parameters (that requires the computation of the partition function), the same group proposed the Pseudo Conditional Random Fields (PCRFs) [Lee 2008]. A logistic regression classifier's output is smoothed using two parameter free pairwise potential functions (one for label interaction and the second for voxel interaction). The learning task is thus reduced to learning the logistic regression's parameter. Inference was performed using graph cuts. The system was tested for multispectral (T1, T1CE, T2) 2D segmentation of 11 contrast enhancing tumors and showed similar results with their previous CRF formulation with the benefit of substantial decrease on the learning time.

Bauer et al. [Bauer 2011] also coupled SVM classification with CRF smoothing. They use a two stage smoothing, first a 3D regularization that is orientation dependent (weaker in the z axis due to anisotropic data) to separate tumor from background and second a stronger 2D regularization on a highly connected neighborhood (8 neighbors in 2D) to perform a 6 classes classification (CSF, White Matter, Grey Matter, Necrosis, Active Tumor and Edema). They exploited recent advances in MRF/CRF optimization using the Fast-PD algorithm, yielding a very fast algorithm. The method was tested on 10 brain tumor patients using multi spectral data (T1, T1CE, T2 and FLAIR).

To conclude, graph based segmentation methods are attractive due to their ability to recover good optima as well as their modularity and scalability. Markov Random Field models in particular enable to couple inter node interaction with individual node classification properties. Further-

more, they are computationally efficient thanks to the recent advances in discrete MRF optimization techniques. Although prior knowledge can easily be introduced in an MRF formulation, current methods are limited to local priors that model the dependencies between neighboring nodes. Hierarchical models and position dependent smoothing (CRFs) offer better flexibility and stronger knowledge but still lack global structural information on the brain.

2.5 Atlas based Segmentation

Atlas based segmentation methods have gained a lot of attention in the recent years for automatic segmentation of medical images mostly due to the progress made in the field of deformable registration [Sotiras 2013]. They allow the introduction of prior anatomical knowledge in the segmentation process by registering (i.e. align) an annotated anatomical template or a probabilistic atlas to the volume to be segmented (target volume). The method is particularly interesting for the segmentation of brains due to their relatively limited inter patient variability compared to other organs. When the structures to be segmented are present both in the atlas and target volume, the task of segmentation is converted into a registration task (the segmentation by registration method) [Bondiau 2005]. A different approach has to be adopted when the object to segment is a tumor that is not present in the atlas.

This section will first introduce the problem of medical image registration that is the basis of atlas construction and atlas based segmentation, followed by a review of the different reference atlases used for brain segmentation and their construction. Eventually, the existing applications of atlas registration to tumor segmentation will be reviewed. A more detailed review of some key deformable registration methods, applied to the context of brain tumors, will be presented in chapter 4.

2.5.1 Medical Image Registration

Medical Image Registration is, concurrently to image segmentation, among the most active research areas in the domain. It is defined as the process of aligning two (or more) images by establishing point to point correspondences. Given a floating image J and a fixed image I defined on a domain Ω, the aim is to find the transformation $\mathcal{T}$ that will map the floating image to the fixed image:

$$I(\mathbf{x}) = J \circ \mathcal{T}(\mathbf{x}) \tag{2.29}$$

We differentiate monomodal (same image modality) and multimodal (two different modalities) registration, as well as intra-patient (the two images belong to the same patient) and inter-patient registration. Intra-patient monomodal registration is used for temporal follow up. The objective is to compare the patient's image taken at different times and evaluate changes, particularly the evolution of a pathology with or without treatment. Registration can be necessary due to the patient's motion or changes in the image acquisition process. The different imaging modalities display complementary informations but are obtained in different reference systems. Aligning multimodal

images renders direct comparison possible such as the association of structural with functional information. Inter-patient registration refers to atlas construction and alignment to a new patient for segmentation or population studies.

The registration process is typically carried out by minimizing a similarity criterion that evaluates the differences between the images to match:

$$\mathcal{T} = \underset{\mathcal{T}}{\operatorname{argmin}}\, Sim(I, J \circ \mathcal{T}) \tag{2.30}$$

The problem is characterized by 3 key elements: the image features to match and corresponding matching criteria $Sim(.,.)$, the transformation model that describes how the image will be deformed, and the optimization technique used to recover the optimal transformation.

2.5.1.1 Matching Criteria

When trying to align two images, we rely on the concept that both images have the same structures and try to find their correspondences. Two kinds of approaches can be considered to match the structures: geometric and iconic registration. Geometric approaches define a set of key points in the image corresponding to structural or geometrical landmarks, or extract surfaces from the pre segmented structures' boundaries. The matching criteria to be minimized is defined as the distance between corresponding points or surfaces in both images. Geometric methods have the advantage of relying on the anatomical structures and yield transformations that are more faithful and preserving of the anatomy. On the other hand, the registration is as good as the landmarks whose automatic selection is not trivial.

Iconic (or intensity based) approaches seek to optimize a similarity measure based on the voxels' intensity values. The measure is dependent on the intensity relationship between the two images The underlying principle is that when the images are matched, their intensities should verify a certain relationship. The simplest relationship is the identity. The images' intensities are assumed to have the same or similar intensities in corresponding structures. This relationship is considered when dealing with monomodal images and allows direct comparisons of the images intensities. The most popular measures are the Sum of Absolute Differences (SAD) and the Sum of Squared Differences (SSD).

A more sophisticated approach for monomodal registration is to assume an affine relationship between the images intensities to acknowledge the potential relative intensity changes between the images to be registered. In this setting, the criterion of choice is the correlation coefficient [Brown 1992, Kim 2004]:

$$CC = \frac{Cov(I, J \circ \mathcal{T})}{\sqrt{Var(I)}\sqrt{Var(J \circ \mathcal{T})}} \tag{2.31}$$

Those monomodal similarity criteria directly compare the intensities of the two images. However, as it is the case for image classification, intensity alone is not that discriminative and can lead to ambiguous matchings and convergence to bad local minima solutions. This issue has motivated the work of Shen et al. [Shen 2002], where they proposed to use feature or attribute vectors that

encompass neighborhood information and local anatomical information based on pre established tissue segmentations.

The definition of a matching criterion becomes more difficult when dealing with multimodal images. Given the impossibility to estimate the relationship between multimodal intensities, it was suggested in [Roche 1998] to assume an unknown functional relationship $I \approx f(J \circ \mathcal{T})$. Given this functional assumption, they proposed to use the Correlation Ratio as a similarity measure:

$$CR = \frac{Var(E(I|J \circ \mathcal{T}))}{Var(I)} \tag{2.32}$$

Information Theoretic approaches assume no existing relationship between the intensities, making them more general and very popular. Since its introduction, the Mutual Information (MI) [Wells III 1996a, Viola 1997, Collignon 1995, Maes 1997] has been widely used (see [Pluim 2003] for a review of MI based registration methods) and is probably the most commonly used multimodal registration criterion. The objective of this measure is to maximize the mutual information of two images which is a combination of their joint and individual entropies. The Shannon entropy of a distribution p can be interpreted as a measure of the distribution's dispersion and is defined as:

$$H(x) = -\sum_{x} p(x) log(p(x)) \tag{2.33}$$

The joint probability distribution $p(I, J \circ \mathcal{T})$ of two images is estimated from the normalized joint histogram of the images intensity values, while the marginal distributions $p(I), p(J \circ \mathcal{T})$ are obtained by summing over the rows and columns of the joint histogram. When the images are aligned, the joint histogram will be sharp and focused, while its dispersion augments as they get misaligned. As a result, minimizing the joint entropy corresponds to aligning the images. This is the idea behind the mutual information defined as:

$$MI(I, J \circ \mathcal{T}) = H(I) + H(J \circ \mathcal{T}) - H(I, J \circ \mathcal{T}) \tag{2.34}$$

Given this definition, the term $-H(I, J \circ \mathcal{T})$ implies that maximizing the mutual information is related to minimizing the joint entropy. The presence of the individual entropies makes it more robust than direct minimization of the joint entropy: transformations that reduce the amount of information in an image are penalized. The Normalized Mutual Information (NMI) was later introduced as a measure less sensitive to the amount of overlap between the images:

$$NMI(I, J \circ \mathcal{T}) = \frac{H(I) + H(J \circ \mathcal{T})}{H(I, J \circ \mathcal{T})} \tag{2.35}$$

Multimodal registration remains a difficult problem that is actively studied. Aside from information theoretic methods, a second kind of approach consists of transforming the problem into the simpler monomodal problem. This can be achieved by simulating one modality into the other [Roche 2001, Wein 2008], or mapping both images into a new space where they are directly comparable [Maintz 2001, Lee 2009, Bronstein 2010].

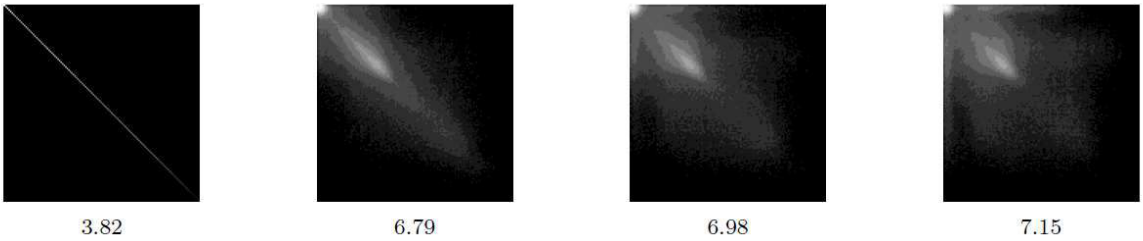


Figure 2.11: *Evolution of the joint histogram between an image and iself as one image is progressively rotated (0,2, 5 and 10 degrees). Below the histogram are the joint entropy values. Example taken from [Pluim 2003]*

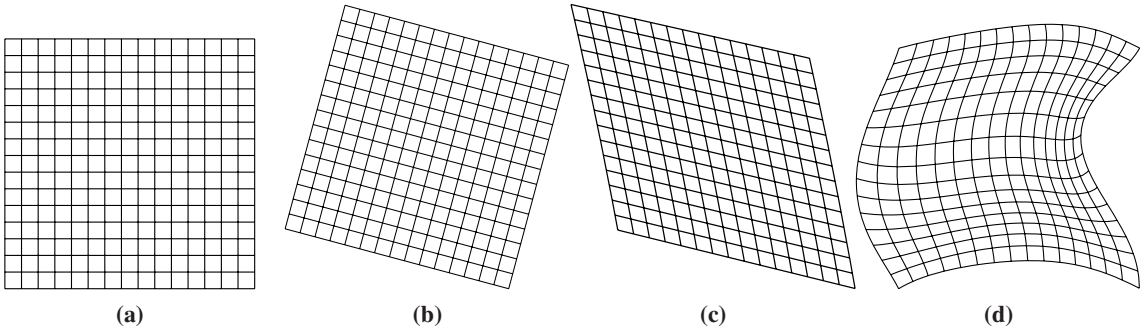


Figure 2.12: *Different types of transformation models. (a) original image, (b) rigid transformation, (c) affine transformation, (d) non-linear transformation.*

2.5.1.2 Transformation Model

The transformation model describes how the image will be deformed and the amount of deformation allowed. The computational complexity as well as the precision of the deformation are dictated by the transformation model. The rigid transformation is the simplest transformation model. It only allows rotation and translation that, in the 3D case, amounts to 6 parameters (degrees of freedom) that need to be estimated. Similarity/Affine registration allow scaling and shearing additionally to translation and rotation (parallel lines are preserved), which corresponds to 12 degrees of freedom in 3D. Rigid and affine transformations are essentially used for same subject registration of organs that display limited movements or deformation in time such as the healthy brain. When stronger deformations are observed, they are often used as preprocessing prior to a dense registration.

If no constraints are imposed on the transformation, the registration problem is referred to as non-linear, dense, curved or deformable registration. The problem is ill-posed with unstable and non smooth solutions. To tackle this issue, a priori knowledge is introduced in the minimization

problem in the form of a regularization term:

$$E(\mathcal{T}(\mathbf{x}) = Sim(F, G \circ \mathcal{T}(\mathbf{x})) + \lambda Reg(\mathcal{T}(\mathbf{x})) \tag{2.36}$$

The parameter λ is a constant that weights the influence of the regularization over the similarity measure. The regularization term ensures smoothness of the deformation field and preservation of the anatomical properties. Two kinds of deformation models can be considered in this setting: parametric transformations and physics based transformations. In both cases, the transformation is usually defined at each pixel position as an additive displacement: $\mathcal{T}(\mathbf{x}) = \mathbf{x} + u(\mathbf{x})$.

Parametric models represent the transformation as a linear combination of basis functions. Those methods are interpolation based: the transformation is calculated on an ensemble of control points that are then used to produce a dense interpolated flow for the entire image. Their interest reside in the fact that the model's complexity (correlated to the number of control points) can be considerably reduced while the richness of the transformation is preserved. The most popular parametric transformation model is the Free Form Deformation model (FFD) [Sederberg 1986, Rueckert 1999]. Here, the control points are placed uniformly across the image with uniform spacing δ and the basis functions used are cubic B-splines. The deformation field is computed as:

$$u(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \mathbf{d}_{i+l,j+m,k+n} \tag{2.37}$$

where $i = [x/\delta_x] - 1$, $j = [y/\delta_y] - 1$, $k = [z/\delta_z] - 1$, $u = x/\delta_x - [x/\delta_x]$, $v = y/\delta_y - [y/\delta_y]$ and $w = z/\delta_z - [z/\delta_z]$. B_l represents the lth basis function of the B-spline and $\mathbf{d}_{i+l,j+m,k+n}$ is the displacement at the corresponding control point. It has been proven [Choi 2000] that the obtained transformation is diffeomorphic if the maximum displacement is limited to 0.4δ.

Physics based deformation models rely on a physical process guided by a force that is associated to the image matching criterion. The deformation model is described by a partial differential equation that is associated to the kind of physical model employed. Among the different physical models used for deformable registration, we can cite elastic registration (stretching of an elastic material) [Bajcsy 1989], viscous fluid models [Christensen 1994] and more famously diffusion models governed by:

$$\Delta u + \mathbf{F} = 0 \tag{2.38}$$

The use of registration through diffusion models was introduced in [Thirion 1998] as an analogy to a thermodynamic concept (Maxwell's demons). Demons are mathematical entities located at the boundary of an object in the image that is modeled as a membrane. Demons apply forces to push the deformable model in the normal direction of the membrane according to local image characteristics. For instance, a demon can be placed at each location in the image where it is not constant. In this case, the corresponding membrane is an isosurface. The demons forces are inspired from optical flow equations but are normalized in order to handle instabilities due to small images gradients. This yields the displacement field gradient or velocity as:

$$v = \frac{(I - J \circ \mathcal{T})\nabla I}{\|\nabla I\|^2 + \alpha(I - J \circ \mathcal{T})^2} \tag{2.39}$$

The method is iterative and alternates between computation of the forces with regularization by Gaussian smoothing. This resulted in a computationally efficient algorithm that lacked a proper theoretical formalism. The first successful attempt to provide theoretical grounds of support for demons can be found in [Pennec 1999], casting the problem as a gradient descent based energy minimization process. In order to recover the algorithm's efficient two step optimization process, it was suggested in [Cachier 2003] to introduce hidden variables that decouple the problem in a matching and regularization step. An extension of Thirion's algorithm to diffeomorphism using this approach was later introduced in [Vercauteren 2008].

Another physics based registration approach that has proven relevant in the presence of brain tumors is the use of biomechanical models to describe the transformation. Such models describe the elastic deformation properties of the different tissues using Finite Element Methods. The main difficulty resides in finding the appropriate model for the anatomical tissues which is far from trivial. It should also be noted that inter subject registration (and by extension, population based atlas construction) is not possible since the deformations from one subject to the next are not natural. Registration methods in the presence of tumors using biomechanical models will be reviewed in chapter 4.

2.5.1.3 Optimization procedure

The optimization procedure is closely related to the design of the transformation model and similarity metric. Continuous methods seek real valued registration parameters Θ. The optimal set of parameters is determined in an iterative fashion given an initial estimate as:

$$\Theta^{(t+1)} = \Theta^{(t)} + \alpha^{(t)}\mathbf{d}^{(t)} \tag{2.40}$$

$\mathbf{d}^{(t)}$ is the search direction at iteration t, and $\alpha^{(t)}$ is the step size along the search direction. Plethora of optimization methods have been proposed that differ in the definition of $\alpha^{(t)}$ and $\mathbf{d}$. Notably, a comparative study of continuous registration strategies can be found in [Klein 2007]. The most famous and simple is the Gradient Descent method that defines the search direction as the derivative of the energy evaluated given the currently estimated parameters. More elaborated strategies involve a combination of the gradient and search direction with the ones at the previous iteration (Conjugate Gradient methods) [Hestenes 1952], computing an approximation of the gradient (Stochastic Gradient Descent) [Kushner 2003] or using higher order derivatives in the search direction (Quasi-Newton methods) [Dennis 1977]. Continuous methods are tuned to the specific registration method and similarity measure for which they are used.

The use of discrete optimization methods has gained a lot of attention in the past years. Contrarily to continuous methods, they perform a global search which makes them robust to initialization while benefiting from fast convergence rates. Another strong advantage of using such an approach is the modularity with respect to the similarity criterion and regularization term. The use of discrete methods is associated to the modeling of the registration problem as a discrete pairwise MRF. Given a predefined discrete displacement set, the registration problem is reformulated as a labeling problem where each node (control point given a parametric transformation model) is assigned

a label that corresponds to a displacement. The MRF energy is defined so that the unary terms correspond to the similarity criterion, while the pairwise term plays the part of the regularization. The optimization problem then boils down to an MRF optimization. Among the many MRF optimization methods developed, Fast-PD and TRW-S have proven successful in the context of image registration [Shekhovtsov 2008, Glocker 2008a]. The main limitation of discrete methods is the trade off between a precise sampling of the search space and control points and computational efficiency. The issue will be addressed in chapter 4.

2.5.2 Brain Atlases

The investigation of the brain's structural and functional organization has motivated the construction of numerous maps of the brain. Brain atlases have been built to create standardized reference coordinate systems in which individual brain maps can be mapped and compared. One of the first reference system (the stereotaxic or Talairach coordinates) was proposed by [Talairach 1988] and became the standard reference in brain mapping. The stereotaxic coordinates are defined based on two landmarks in the brain: the Anterior (origin of the coordinates) and Posterior commissures (AC and PC respectively). They were defined in order to horizontally orient the line connecting the two landmarks and vertically orient the interhemispheric plane. Alongside with the definition of the atlas, a piecewise affine transformation method (the brain is divided into 6 associated blocks) has been proposed to register a volume to the Talairach atlas. Despite the success of the Talairach atlas, it suffered from severe limitations. It was constructed from the single post mortem brain of a 60 year old French woman that was far from representative of the population, had a variable slice thickness and inconsistent data from orthogonal planes. The developments in medical imaging paved the way for the construction of more representative and detailed atlases. Among them, a 3D T1-weighted MR based atlas (the Harvard Brain Atlas) was constructed by the Surgical Planning Laboratory, using the scan of a 25 years old healthy subject that was enhanced by anisotropic filtering [Kikinis 1996]. It was hand-labeled into 150 different structures. To cope with the lack of contrast and high signal to noise ratio, The Montreal Neurological Institute (MNI) constructed a single subject atlas by averaging 27 high resolution T1-weighted scans of the same individual. [Holmes 1998]

Despite offering precise structural information, single subject atlases are not always representative of the population and can introduce a bias. This has motivated the construction of population atlases that are based on an ensemble of scans that are registered and averaged. The reference population atlases have been constructed by the MNI and Laboratory of Neuro Imaging of UCLA (members of the International Consortium for Brain Mapping (ICBM) project). Their first population atlas was the MNI-305 [Evans 1993, Collins 1994], constructed from 239 males and 66 females around 23 years old. All images where registered to the Talairach space and their intensities were normalized. The atlas was a voxel wise average of the intensities over all volumes. Two improved atlases where constructed using the same procedure with 152 (MNI-ICBM152) [Fonov 2009, Fonov 2011] and 452 (ICBM 452) [Mazziotta 2001] subjects. The resulting volumes are blurred where structures vary among subjects and are therefore not suited for structural stud-

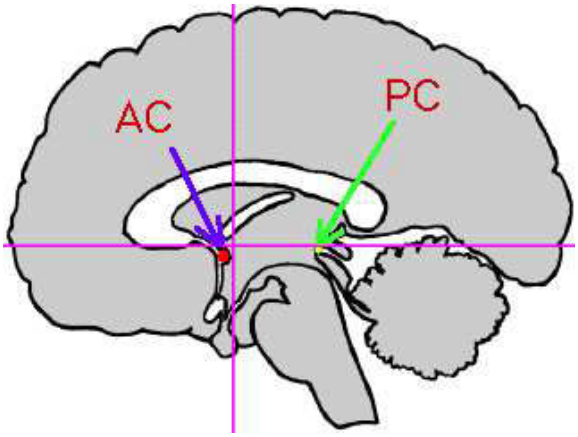


Figure 2.13: *Schematic illustrating the AC-PC line and axes of the Talairach space. Example taken from http://airto.ccn.ucla.edu/.*

ies but offer good templates for registration with limited bias. Similar atlases where constructed for T2-weighted, PD-weighted and Diffusion weighted images. Last but not least, probabilistic interpretation of those atlases were constructed by segmenting the different subjects tissues (white matter, grey matter and CSF) after registration. The probabilistic maps are constructed by averaging the labels over all the subjects. Those atlases describe the spatial distribution of tissue classes over large populations and can be exploited as global anatomical priors. Extended reviews of brain mapping and atlases can be found in [Toga 2001, Cabezas 2011].

2.5.3 Atlas based Tumor Segmentation

The segmentation of brain tumors cannot be mapped into a registration process due to the absence of a tumor in the atlas to be registered. Existing methods exploit the information brought by the atlas as a spatial prior on healthy brain structures. Kaus et al. [Kaus 1999, Kaus 2001] extended an atlas based segmentation method (the ATM SVC algorithm) [Warfield 2000] for meningioma and low-grade glioma segmentation. The method alternates between supervised kNN classification and registration of a manually segmented anatomical template (single subject) based on the current segmentation. The atlas is initially rigidly registered to the volume to be segmented. The features for kNN classification are the voxel intensities as well as spatial features obtained from the aligned atlas. The spatial features are computed by converting each of the tissue classes into a distance map. Using distance maps allows to have a non optimal registration result since it doesn't rely on hard boundaries but on the distance from where the labels are quite certain. The classification is then followed by an elastic deformable registration on the basis of the segmented image [Dengler 1988]. During the registration steps, the tumor is labeled as healthy tissue. A spatial template for tumor is defined at the next iteration based on the tumor segmentation obtained at the previous iteration. The different brain structures (background, skin, brain, ventricles and tumor) were segmented hier-

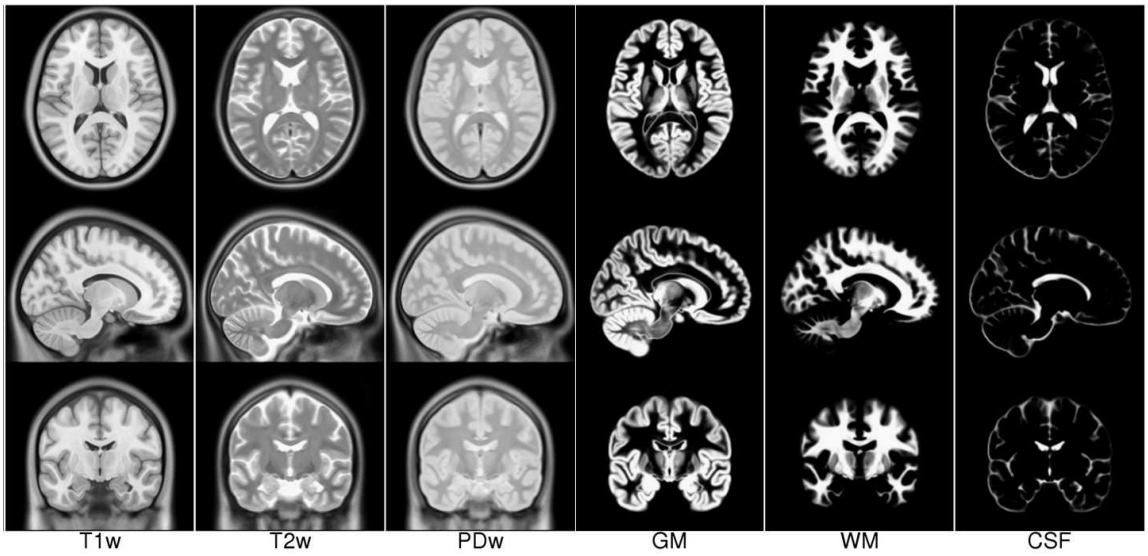


Figure 2.14: *MNI-ICBM-152 probabilistic atlas. From left to right: T1, T2 and Proton Density weighted images ; Gray Matter, White Matter and CSF probability maps. Figure extracted from [Fonov 2011]*

archically. The method was tested on 20 T1CE images: 7 meningiomas and 14 low grade gliomas. The method inherits the drawbacks of kNN classification and implies strong homogeneity of the tumor appearance and limited spectral overlap with the healthy tissue. The successive applications of registration and classification also make the algorithm slow.

Cobzas et al. [Cobzas 2007] integrated atlas priors in a combined statistical learning/level set framework. The level set is defined in a region competition manner and evolves based on pre estimated probabilities. After registration, the atlas prior was integrated in a high dimensional feature set for supervised statistical learning of the two classes probabilities.

The classification decisions were smoothed by an MRF prior designed as a Potts model. The optimal parameters (distributions mean and covariance, bias field parameters and MRF parameters) and tissue classes were estimated using the EM algorithm. Automation of the method was achieved using a probabilistic brain atlas (ICBM152). After affine registration of the atlas to the subject, the probabilities were used as a rough prior estimation of the classification at every voxel. Furthermore, the atlas probabilities values were used as a spatial prior to increase the quality of the classification.[Moon 2002, Prastawa 2003] built on this work and adapted it to the presence of tumor and edema. For each patient, the probabilistic atlas is modified to include prior probabilities on the tumor and edema positions. The tumor prior probability map is constructed based on contrast enhancement. A difference image between T1 and T1CE images is constructed. A mixture of distributions is fitted to the difference image's histogram: two Gaussian distributions that correspond to noise and errors in registration and a Gamma distribution for contrast enhancement.

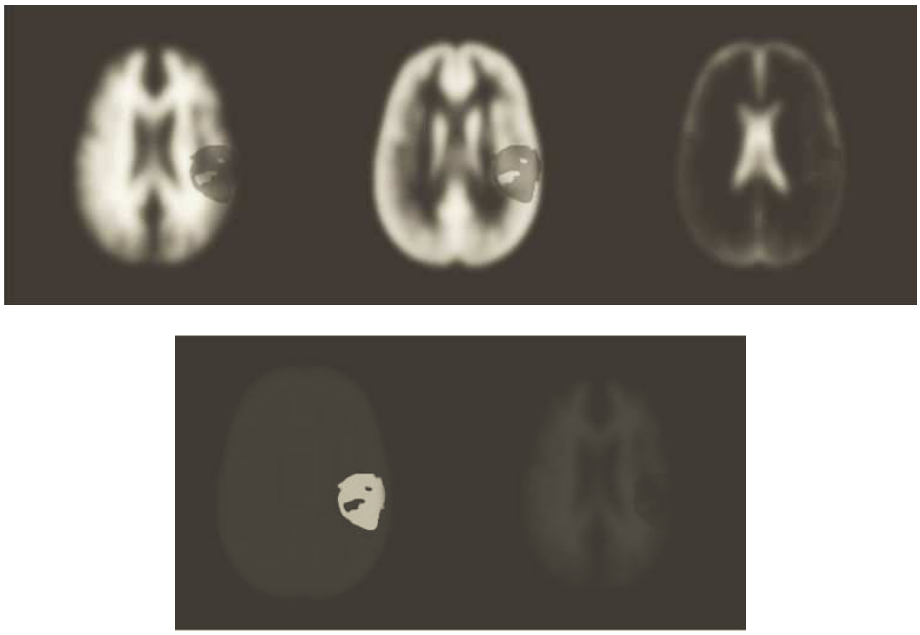


Figure 2.15: *Modified MNI-305 atlas for integration of tumor and edema probability [Prastawa 2003]. Top row: healthy probability (White Matter, Grey Matter, CSF). Bottom row: tumor(left) and edema (right) probabilities.*

The probability that corresponds to contrast enhancement is used to convert the difference image into a tumor prior probability.To remove contrast enhanced areas that do not correspond to the tumor, a smoothing regularization process is included using the region competition level set method developed by [Ho 2002].The edema prior class is introduced similarly as a fraction (20 %) of the white matter prior probability. The method was applied to 5 glioblastoma and meningioma cases and used T1 and T2 images for classification.

Another approach, similar to the work of [Gering 2002], was to detect the tumor voxels as outliers with respect to the normal voxels. The concept was proposed by Van Leemput et al. [Van Leemput 2001] as an extension of their previous work by including an outlier class for multiple sclerosis segmentation. Each class probability is here defined as a mixture of a normal distribution and a uniform distribution δ corresponding to the outlier class:

$$p_k(I(\mathbf{x_i})|\Phi_i) = (1-\epsilon)\mathcal{N}(\mu_k, \Sigma_k) + \epsilon\delta \tag{2.41}$$

Where $\epsilon \in [0, 1]$ represents the portion of outliers. After computation of the parameters and class memberships, a voxel will be assigned to the outlier class if its membership score is below a threshold for all classes. The intensity model employed here for multiple sclerosis is too simple for a direct application to tumor segmentation.

The same idea (model lesions as outliers) guided the work of Prastawa et al. [Prastawa 2004].

They presented a more general framework in comparison to their previous work [Prastawa 2003] that was limited to enhanced tumors featuring edema. The method consists of three steps: first the tumor and edema are detected as outliers from the healthy classes given their intensity profile, then the presence of an edema is detected while the last step refines the segmentation output using spatial and geometric constraints. Affine registration of a probabilistic atlas (ICBM atlas) allows the random selection of a set of training samples representative of each healthy class (high atlas probability for White Matter, Grey Matter and CSF) but that are contaminated with tumor and edema samples (the abnormal class). The healthy multimodal (T1 and T2) intensity distributions are assumed to be Gaussian. The minimum covariance determinant [Rousseeuw 1999] is used to estimate the classes mean and covariances due to its robustness to the presence of outliers. The abnormal class is then formed by rejecting samples that are three standard deviation away from their class distribution. Once the samples are correctly classified, the problem is defined as a MAP problem. The likelihoods are estimated using Parzen kernel density estimation [Duda 2001]:

$$P(I(\mathbf{x})|C) = \frac{1}{N}\sum_{i=1}^{N} K_\lambda \left(I(\mathbf{x}) - I(\mathbf{p_i})\right) \tag{2.42}$$

where K_λ is the Gaussian kernel, λ is its standard deviation or kernel bandwidth that determines the amount of smoothing on the estimated probability and $\mathbf{p_i}$ is a training sample. The spatial priors for healthy classes are extracted from the atlas, and a fraction of the sum of the white and grey matter probabilities are used as prior for the abnormal class. In an EM style approach, this classification process was jointly considered with the bias field correction method of [Van Leemput 1999]. Tumor and edema classes are separated by clustering based on their T2 intensity and spatial constraints are enforced using a region competition level set.

Menze et al. [Menze 2010] proposed an algorithm for tumor segmentation that strongly exploits the fact that the tumor has different visibility in multi-channel images. They propose a generative model in which each modality is modeled and segmented independently. Each class (including the tumor class) and modality is modeled as a Gaussian distribution combined with a smoothing MRF prior. A vector of tumor appearance describes the observation of a tumor at each voxel for each modality. An EM style approach is adopted to alternatively estimate the classification probabilities from which are derived the multi channel classifications, and the Gaussian distribution's parameters. The global segmentation is obtained by averaging the classification decisions over all channels. The method is initialized by segmenting the image into 4 classes (White Matter, Grey Matter, CSF and outlier) using the method of Van Leemput et al. [Van Leemput 2001]. The outlier class is defined as being more than three standard deviations away from the centroid of any normal class. The method was tested on 25 glioma patients using T1, T1CE, T2 and FLAIR volumes.

Atlas based tumor segmentation methods have a strong dependency on the quality of the registration. Despite numerous efforts to make the segmentation robust to errors in registration, the methods will fail if the brain is deformed by the tumor. Detection as an outlier also implies that the tumor's appearance varies significantly from the one of the healthy tissue, an assumption that

does not always hold given the frequent intensity overlaps. Nevertheless, the use of anatomical and probabilistic atlases is a powerful tool for incorporation of prior spatial information on the expected localization of the tumors.

2.6 Summary

Despite an extensive amount of work dedicated to brain tumor segmentation, the problem remains difficult and segmentation results are often not satisfactory. Generalization of the existing methods to several kinds of tumors (or pathologies) is often difficult. Furthermore, few methods have been adapted to heterogeneous DLGG segmentation. The focus of early methods was on contrast enhanced and homogeneous tumors and while glioblastomas have been mostly studied in the recent years. Additionally, many methods are based on the use of different image modalities which are not systematically available in a clinical setting. Despite the need of a prior training phase, machine learning methods offer a flexible and efficient way of capturing the tumor's properties that are not limited to intensity values. Coupled with MRF models, the i.i.d assumption is relaxed and smooth solutions can be obtained very fast due to recent advances in MRF optimization techniques. An additional advantage of MRF modeling is the natural inclusion of prior knowledge in the formulation. Such prior knowledge can be obtained via the use of anatomical or probabilistic atlases.

Next chapter will present a DLGG segmentation method that introduces prior atlas based knowledge in an MRF framework. Instead of detecting the tumor as an outlier with respect to the healthy tissue distribution, a tumor specific atlas is constructed describing the most likely position of tumor voxels.

Chapter 3

A Probabilistic Atlas of Diffuse WHO Grade II Gliomas' Preferential Locations

3.1 Introduction

Brain atlases are excellent tools for the study of the brain's anatomy and the variations within populations. Futhermore, they offer a reference template for registration and atlas based segmentation tasks. They are however, not representative of diseased populations since the pathology tends to alter or deform the brain's structures. Several atlases are available with repositories of annotated images for a diversity of pathologies. Such images are a rich source of information but not in the same coordinate space and therefore not directly comparable.

This has motivated the construction of disease specific atlases that reflect the particular clinical subpopulation affected by the disease and how it alters the brain. Such atlases enable to study the evolution and origin of the disease as well as the variations from one patient to another. Aside from aligning all volumes in the same coordinates system, the difficulty resides in obtaining a statistical interpretation of the inter subject variability. The construction of disease specific atlases has notably been one of the focuses of UCLA's Laboratory of Neuro Imaging. Among notable disease specific atlases, Thompson et al. [Thompson 2000] have constructed atlases dedicated to the impact of schizophrenia and Alzheimer's disease on the brain. Maps of anatomical variability are constructed by comparing the deformation fields obtained by deformable registration of all volumes.

In the context of Diffuse Low Grade Gliomas (DLGG), the construction of an atlas mapping the tumors appearances in the brain was motivated by the fact that the management of brain tumors is highly dependent on their location. The impact of the tumor is dependent on the eloquence of the region in which it is. Several studies have reported a functional reorganization of the brain induced by the slow growing tumors [Desmurget 2007, Duffau 2005] while it was inferred in [Ius 2011] that

certain areas have low compensatory abilities. These observations are directly linked to surgery planning for tumor resection, where the eloquence of the region and its compensatory ability can determine the safety of the surgery and amount of resection possible. Furthermore, tumors could have a location dependent behavior due to local differences in their molecular biology [Gozé 2009, Ren 2012].

An interesting observation was made in [Duffau 2004], inferring that DLGGs tend to appear in preferential locations. This assumption was explored for all gliomas in [Larjavaara 2007, Gooya 2012] where it was observed that the distribution of gliomas in the brain is non-uniform. Following this observation, this chapter introduces a probabilistic atlas that maps the DLGGs appearances in the brain and provides their preferential locations. This is carried out via the construction of a complete graph, where each node represents a tumor and the edges represent the spatial proximity between them. The tumors' preferential locations are then identified by extraction of a sparse subgraph, where each cluster corresponds to a preferential location. This results in a DLGG atlas where the central position (cluster center) and spatial extension (position of tumor voxels in the cluster) of each preferential location is represented. An application of this atlas is proposed in the second part of this chapter for the segmentation and characterization (integration in the atlas) of a new patient's tumor with spatial priors. This is carried out via a two-level Markov Random Field formulation based on position priors inferred from the atlas. The tumor is characterized and segmented based on a prior tumor detection through statistical classification, while the knowledge on the tumors' spatial extension in the clusters is used as a prior knowledge on the position of the voxels of the tumor to be segmented.

The remainder of this chapter is organized as follows: first the methodology for the construction of the low-grade glioma atlas is presented in section 3.2, while section 3.3 presents a direct application of the atlas for tumor segmentation. The experimental validation is part of section 3.4 and discussion concludes the chapter.

3.2 Spatial Position Mapping of Diffuse Low-Grade Gliomas

To construct the atlas, we assume the existence of a database of patients MRIs featuring a DLGG. The position of each tumor is known through binary maps obtained by experts' manual segmentation. To allow position comparison, all volumes are to be aligned to the same reference template of a healthy brain. This section presents the construction of the complete graph (registration and proximity measure between the tumors) followed by the estimation of its optimal clustering.

3.2.1 Graph Construction

3.2.1.1 Registration to a Common Reference Space

The mapping of all annotated volumes to the same reference space could be done through deformable registration. However, the presence of tumors violates the correspondence hypothesis, which can cause displacements errors and a shift in the position of the tumor. Discarding the

pathology during registration was considered in [Parisot 2011] but as will be shown in chapter 4, such an approach can still lead to errors in registration at the vicinity of the tumor. Despite being less precise, affine registration is the most common choice in atlas construction, because it handles properly linear variations and therefore roughly conserves the anatomy so that the form and location of the tumor is preserved and comparable. The use of affine registration reflects the diversity of the anatomy, which makes the atlas more adapted for the integration and tumor segmentation of a new, unsegmented patient.

We adopt the block matching method of [Ourselin 2000] to affinely register our data set to the same reference template using the MedInria software (`http://med.inria.fr/`). Instead of globally minimizing an objective function, the idea of block matching is to perform registration in a sequential manner by minimizing a local objective function between image regions (blocks).

For rigid or affine registration, the registration is performed in a multiscale iterative manner where two successive tasks are carried out at each iteration. First, the displacement field is computed between the images by maximizing a local similarity measure. This is done by moving a block of the floating image I and comparing it to blocks in the reference target image J that have coincident positions. The size of the blocks and extent of the search space are constant at a given scale level. The Normalized Mutual Information similarity metric is used in MedInria.

The block that maximizes the similarity measure is chosen, which yields a correspondence between the centers of the blocks (a_i, b_i). Inadequate displacement fields that correspond to a similarity measure below a given threshold (uniform blocks or absence of correspondences) are not considered in the subsequent step. The transformation $\mathcal{T}$ that best fit the obtained displacement field is computed by least trimmed square regression:

$$\begin{aligned} &\underset{\mathcal{T}}{\operatorname{argmin}} \sum_{i=1}^{h} \|a_i - \mathcal{T}(b_i)\|^2 \\ &h = \left\lfloor \frac{n}{2} \right\rfloor \end{aligned} \tag{3.1}$$

where n is the number of displacement vectors. An illustration of the block matching search process is shown in Fig. [3.1]

All volumes in the data set are registered to the same reference template that is an arbitrarily chosen healthy subject, following this scheme. The obtained displacement fields are then applied to the segmentation binary maps. The next step consists in measuring the proximity between the different aligned tumor segmentation maps.

3.2.1.2 Proximity Between Tumors

When measuring the proximity between tumors, there are two issues to be considered: (i) the registration errors and (ii) the different sizes of the tumors. Here, we are not interested in the size and extent of the tumor, but solely on its position in the brain reference space. Measuring the distance between tumors as surfaces or point clouds could penalize small tumors that would be further away from its neighbors than a large tumor. Furthermore, considering surfaces makes

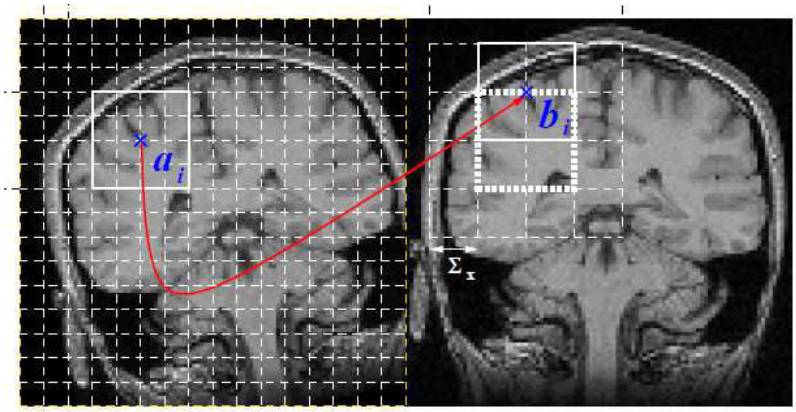


Figure 3.1: *Illustration of the block matching process. The floating image is on the left and target on the right. Figure extracted from [Ourselin 2000]*

the distance measure more sensitive to the registration and miscorrespondences that are bound to happen with the simple affine registration scheme. This has motivated the use of the tumors' center of gravity as a reference position, computed by averaging all tumor pixels coordinates. The proximity between the tumors is then computed as the Euclidean distance between the centers [1].

The distance is computed between each pair of images. This constitutes an adjacency matrix that is used to construct a complete graph $\mathcal{G}_{LGG} = (\mathcal{V}_{LGG}, \mathcal{E}_{LGG})$ where each node corresponds to a patient/tumor and is spatially associated to the center of gravity of the tumor, while the edges connect all pairs of nodes $(\mathbf{p},\mathbf{q}) \in \mathcal{V}_{LGG}$ and are weighted by the euclidean distance between tumors centers $d(\mathbf{p}, \mathbf{q})$. The obtained complete graph can be observed in Fig.[3.2], where the color and thickness of the edges represent their strength (from red and thick to blue and thin). From this complete representation, the existence of high and low density areas for tumors can be observed, most notably around the temporal lobes.

3.2.2 Graph Clustering

The next step towards identifying the DLGGs' preferential locations is to regroup nearby nodes based on their proximity. This can be done by performing clustering on the graph. The concept of clustering, as introduced in section 2.4, is to regroup the data into segments so that samples within the same cluster are similar to each other and dissimilar with respect to the elements associated with other clusters. In this section, we will only consider "hard" clustering, where each sample is assigned to a single cluster. In our context, nodes in the same cluster should be close and present in the same anatomical area. The clusters made of the highest number of nodes would

[1] The aforementioned arguments are also supported from experimental evidence. Different distance metrics have been used, including for example the Mahalanobis distance between tumors. The distance between the tumors' centers of gravity has been proven to be more robust.

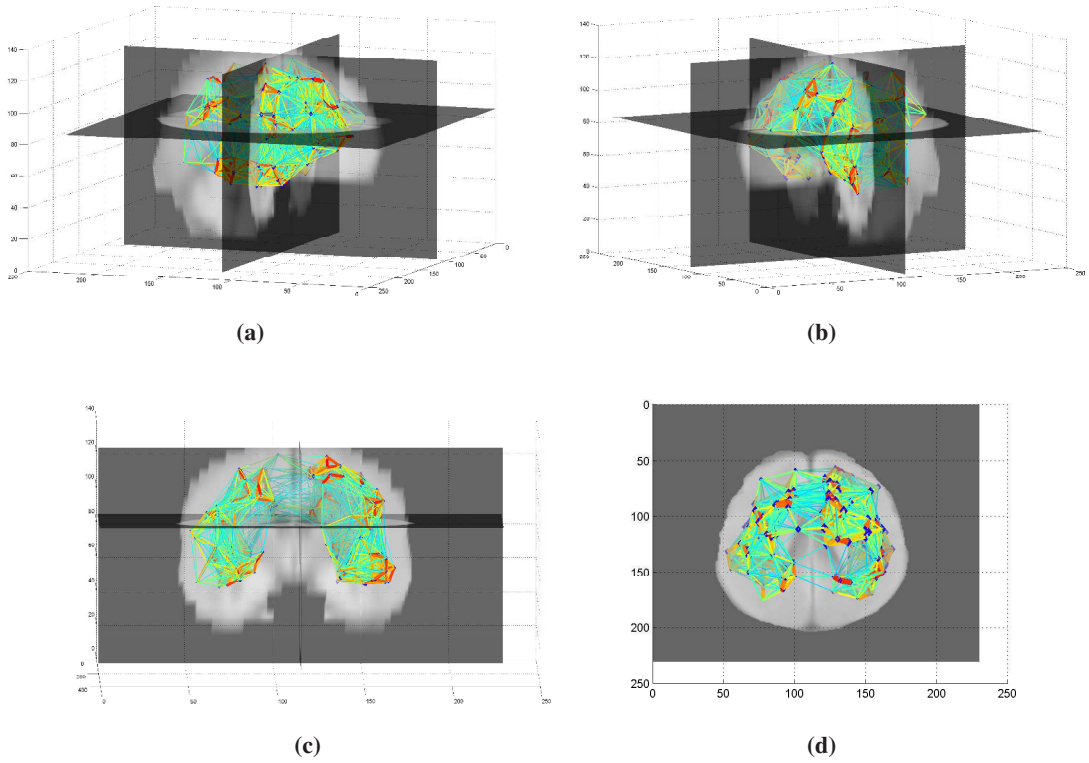


Figure 3.2: *Visualization of the complete graph superimposed to the mean registered image. For visibility reasons, edges of weight larger than 35 are not shown. Short distances appear red and thick while large distance values are thin and blue.*

then correspond to the preferential locations, while an absence of cluster in a region suggests an extremely rare location for tumor appearance. Obtaining compact and well separated clusters, as well as cluster free areas would confirm the existence of preferential locations.

Clustering has been a well studied problem over the past decades, in a number of fields. Early popular clustering algorithms refer to hierarchical methods [Johnson 1967]. Hierarchical or agglomerative clustering initially considers all samples as clusters. The clusters are then progressively obtained by merging neighboring clusters together. The process continues until there is only one cluster left. This results in a dendogram, a tree that represents the progressive clusters merging. The clustering is chosen by selecting a number of clusters and cutting the grouping tree accordingly. There are two issues with such a method, the first refers to the absence of "cluster center" representing each population while the second to the choise of the number of clusters being retained. More adapted to our problem, Centroid based clustering methods group the data by selecting cluster centers and assign the nodes to their closest center. Each cluster center is then

representative of the cluster's properties. The most popular centroid based clustering method is the K-means algorithm [MacQueen 1967].

K-means clustering is an instance of the EM algorithm where the objective is to minimize the square distance between the samples and the cluster center:

$$J = \sum_{k=1}^{K} \sum_{\mathbf{p} \in \mathcal{V}_{LGG}} \|\mathbf{p} - \bar{c_k}\|^2 \tag{3.2}$$

Given the expected number of clusters K, and an initialization of the centers $\{\bar{c_1}, .., \bar{c_K}\}$, the K-means algorithm alternates between assigning the samples to the closest center and reevaluating the centers as the mean coordinates of all samples within the cluster. The K-means algorithm is sensitive to initialization which can be a drawback when there is no pre assumption on the position of the clusters. Furthermore, the cluster centers are constructed as the mean value of the samples in the clusters, which can be sensitive to the presence of outliers and not represent the cluster appropriately since there is no guarantee that the mean lives on the same manifold as the one of the samples. These issues are often addressed by considering different initializations both in terms of the number of populations as well as their centers (this is done by randomly sampling the population) and then adopting criteria from information theory to determine the optimal result.

We adopt a more robust method [Komodakis 2008a] that has been shown to outperform the K-means algorithm. The method simultaneously estimates (i) the number of clusters K (ii) the cluster centers $\{\bar{c_1}, .., \bar{c_K}\} \in \mathcal{V}_{LGG}$, and (iii) the remaining nodes' assignments to the different clusters $A_p \in \{C_1, ..., C_K\}$. The objective function aims at minimizing the distance between nodes and their assigned cluster center:

$$\min_{K} \min_{\bar{c_k}} \min_{A_p} \sum_{k=1}^{K} \sum_{\mathbf{p} \in \mathcal{V}_{LGG}} \delta(C_k, A_p) d(\mathbf{p}, \bar{c_k}) + \alpha \sum_{k=1}^{K} V_{pen}(\bar{c_k}) \tag{3.3}$$

where $\delta(.)$ is the Kronecker delta function and α is a constant parameter that balances the importance of the two terms. The first term of the objective function assigns nodes to the cluster of which center they are the closest according to a given distance on the manifold of the samples' space. The second term is introduced in order to avoid the trivial solution of defining each node as a cluster center. This can be seen as a regularization term that measures the *stability* of a node in order to determine which nodes should be selected as centers. In the context of seeking preferential locations, we aim at selecting nodes that are at the center of high density areas. We define the penalty term as the mean distance over all samples in the graph in order to identify such nodes:

$$V_{pen}(\mathbf{p}) = \frac{1}{N} \sum_{\mathbf{q} \in \mathcal{V}_{LGG}} d(\mathbf{p}, \mathbf{q}) \tag{3.4}$$

The idea behind this term is that nodes should be selected as center if they are close to many other nodes, *i.e.* if its distance value with an important set of nodes is low. Contrarily to the K-means

algorithm, it should be noted that the centers are among the samples of the population on which clustering is to be performed and that they are chosen as the ones situated on dense (statistical interpretation) manifold areas. This makes the center selection more robust to outliers and more faithful to the actual cluster structure. The additional advantages of this clustering approach is that it is modular with respect to the distance metric between samples while the optimization also recovers the optimal number of clusters.

Similarly to the Fast-PD algorithm for MRF optimization, the near optimal clustering is recovered using linear programming by minimizing the gap between the primal and dual problems. The problem is reformulated as the following relaxed linear integer program:

$$\text{PRIMAL-IP} \equiv \min_{\mathbf{x}} \sum_i V_{pen}(i)x_i + \sum_{j,i} d(\mathbf{i},\mathbf{j})x_{ji} \tag{3.5}$$

$$\text{s.t.} \sum_i x_{ji} = 1 \tag{3.6}$$

$$x_{ji} \leq x_i \tag{3.7}$$

$$x_{ji}, x_i \in \{0,1\} \tag{3.8}$$

The binary variable x_{ji} indicates whether the sample j has been assigned to node i or not, while the binary variable x_i indicates if node i has been selected as a cluster center. The constraints of equation 3.6 simply express the fact that each observation should be assigned to one central node, while the constraints of equation 3.7 impose that if observation j has been assigned to node i then node i must obviously be chosen as a cluster center. The most important issue for such an integer program optimization is the correct assignment of the variables x_i, i.e., deciding which nodes will be chosen as the clusters centers. The center selection process is guided by the nodes' stability. Once the central nodes are computed then each observation can be trivially assigned to its nearest central node. This process repeats until no more stable nodes can be found. The resulting method does not require any initialization and provides near-optimal clustering results both in terms of number of clusters as well as in terms of central nodes and remaining nodes assignments.

3.2.3 Clustering Selection

The parameter α weighting the cost of introducing new clusters versus the overall assignment distance error plays a critical role in the obtained clustering. We seek to find the clustering that is the most representative of the data. To this end, we rely on cluster validity indices that provide a quantitative measure of how good the obtained result is. Two kinds of validity criteria exist: external and internal indices. The external indices imply prior knowledge about the data and the expected clustering. They are used to compare the obtained clustering structure with the a priori knowledge. In practice, external criteria are mostly used to evaluate a clustering method as the partition is rarely known and internal criteria are the most useful. In this work, we make use of three widely used internal cluster validity measures: the Dunn index [Dunn 1974], the Davies-Bouldin index [Davies 1979] and the Silhouette index [Rousseeuw 1987].

The **Dunn** index is among the simplest clustering quality measures. Its goal is to determine if the clusters are compact and well separated. The index, introduced in [Dunn 1974] is defined as the ratio between the minimum distance between clusters d_{C_i,C_j} and the maximum cluster diameter $diam(C_k)$:

$$D = \min_{i=1,\dots,K} \min_{j=1,\dots,K} \frac{d_{C_i,C_j}}{\max_{k=1,\dots,K} diam(C_k)} \tag{3.9}$$

In [Dunn 1974], the distance inter cluster is defined as the minimum distance between any samples in the two different clusters while the cluster diameter is the maximum distance between two samples of the same cluster:

$$d_{C_i,C_j} = \min_{\mathbf{p}\in C_i, \mathbf{q}\in C_j} d(\mathbf{p},\mathbf{q}) \tag{3.10}$$

$$diam(C_k) = \max_{\mathbf{p},\mathbf{q}\in C_k} d(\mathbf{p},\mathbf{q}) \tag{3.11}$$

Compact and well separated clusters yield a large Dunn index (large distance inter cluster and small cluster diameter). The Dunn index suffers from several shortcomings, the most important being its severe sensitivity to noise and outliers as they can strongly impact the diameter and inter cluster distance. In order to obtain more robust measurements, several Dunn-like indices can be considered with different definitions of the diameter and cluster distance.

A graph theoretic formulation was notably proposed in [Pal 1997]. Each cluster was represented using graphs of known structure (Minimum Spanning Tree (MST), Gabriel Graph or Relative Neighborhood Graph) and the cluster diameter was defined as the maximum edge value in the graph. We will review the MST based Dunn index that is used in this chapter. Given the set of nodes $\mathcal{V}_{C_k}$ that belong to cluster C_k and the corresponding complete graph $\mathcal{G}_{C_K}$ that connects all pairs of nodes, a spanning tree is the minimal subgraph that connects all nodes together. The weight of the spanning tree is defined as the sum of the weights (*i.e.* distances between nodes) of all the edges in the tree. The Minimal Spanning Tree $\mathcal{G}_{C_K}^{MST} = (\mathcal{V}_{C_k}, \mathcal{E}_{C_k}^{MST})$ is the weighted graph that has the smallest weight. Having constructed such a graph per cluster, the diameter is then defined as the edge of the MST that has the maximum weight $diam(C_k) = \max_{\mathbf{p},\mathbf{q}\in\mathcal{E}_{C_k}^{MST}} d(\mathbf{p},\mathbf{q})$. For increased robustness, the inter cluster distance is modified and defined as the distance between the clusters centers $d_{C_i,C_j} = d(\bar{c_i}, \bar{c_j})$.

Similarly to the Dunn index, the **Davies-Bouldin** index aims at identifying compact and well separated clusters. A similarity measure R_{ij} between the clusters C_i and C_j is computed based on a cluster dispersion measure S_i and compared to an inter cluster dissimilarity measure D_{ij}. The similarity measure should verify the following conditions:

$$\begin{aligned}
&R_{ij} > 0 \\
&R_{ij} = R_{ji} \\
&\text{if } S_i = 0 \text{ and} S_j = 0 \text{ then } R_{ij} = 0 \\
&\text{if } S_j > S_k \text{ and} D_{ij} = D_{ik} \text{ then } R_{ij} > R_{ik} \\
&\text{if } S_j = S_k \text{ and} D_{ij} < D_{ik} \text{ then } R_{ij} > R_{ik}
\end{aligned} \tag{3.12}$$

x Usually, the similarity, dispersion and dissimilarity measures are defined as follows:

$$\begin{aligned} S_i &= \frac{1}{|C_i|} \sum_{\mathbf{p} \in C_i} d(\mathbf{p}, \bar{c_i}) \\ D_{ij} &= d(\bar{c_i}, \bar{c_j}) \\ R_{ij} &= \frac{S_i + S_j}{D_{ij}} \end{aligned} \tag{3.13}$$

The dissimilarity measure corresponds therefore to the ratio of the mean intra-cluster distance (distance of all nodes to their cluster center) for clusters C_i and C_j with the inter-cluster distance (distance between cluster centers). The Davies-Bouldin index is then calculated as the average maximum similarity measure over all clusters:

$$DB = \frac{1}{K} \sum_{j=1}^{K} \max_{i=1,\ldots,K, i \neq j} R_{ij} \tag{3.14}$$

The optimal clustering corresponds to a low value of the DB index, *i.e.* to a low similarity between the clusters. This index is more robust than the Dunn index due to the fact that it relies on average values and distances to the center of the cluster.

The last index considered, the **Silhouette** index, takes a different approach. Computed for each sample $\mathbf{p}$, it evaluates how adequate the sample's cluster assignation is via a score of confidence $s(\mathbf{p})$:

$$s(\mathbf{p}) = \frac{b(\mathbf{p}) - a(\mathbf{p})}{\max(a(\mathbf{p}), b(\mathbf{p}))} \tag{3.15}$$

$a(\mathbf{p})$ is the average distance between sample $\mathbf{p}$ and all the remaining elements assigned to the same cluster (dissimilarity to all other samples in the same cluster). For each cluster C_k the node hasn't been assigned to, the same dissimilarity measure $d_{\mathbf{p},C_k}$ is computed between node $\mathbf{p}$ and all the nodes in the cluster. We define $b(\mathbf{p}) = \min_{k=1,\ldots,K} d_{\mathbf{p},C_k}$, which can be seen as the average distance between $\mathbf{p}$ and all the elements in the closest cluster. The Silhouette index takes values between -1 and 1. If the value is close to 1, the node has been assigned correctly. A value close to zero suggests that the node is equally far away from 2 clusters, while a silhouette index close to -1 infers that the node has been misclassified. To evaluate the quality of the clustering, we compute the global silhouette index:

$$GS = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{n_j} \sum_{\mathbf{p} \in C_j} s(\mathbf{p}) \tag{3.16}$$

where n_j is the number of samples in cluster C_j. The highest global silhouette index corresponds to the best clustering, where the individual Silhouette indexes are closest to 1. The silhouette index gets its name from the fact that it offers a visualization tool of the quality of the clustering. The so-called *silhouette* of the cluster can be plotted, displaying which samples are well classified and the relative quality of clusters, narrow silhouettes indicating better clusters.

In order to select the best fitting clustering, the three different indices are computed for several values of α and the parameter that yields the maximum combined index is selected. In the remainder of this chapter, we will consider the optimal clustered graph made of K clusters.

3.3 Tumor Characterization and Segmentation with Spatial Position Priors

Given the possibility to recover properties of the tumors with respect to their localization, it is of high interest to integrate a new, non segmented patient to the atlas and determine which preferential location the tumor belongs to. In parallel, the atlas offers powerful insight on the expected positions and expansion of the tumors and can be used as a spatial prior for segmentation.

This section presents an application of the DLGG atlas for characterization of the tumor's location with respect to the atlas and for segmentation via the use of spatial position priors. For each cluster C_i of the atlas, we build a probability map $P(\mathbf{x}|C_i), i \in [1:K]$ that describes the rate of tumor appearance per voxel $\mathbf{x}$. This is carried out in a similar fashion as the construction of the probabilistic ICBM atlases by averaging the binary segmentation maps of the different tumors in the same cluster. In order to ensure smoothness of the probability maps for clusters with a small number of samples, all probability maps are smoothed by Gaussian filtering. Fig. [3.3] shows examples of probability maps for different clusters, superimposed to the average volume.

We consider an image I defined on a domain Ω featuring a DLGG. An initial tumor detection, performed via statistical classification, is used within a coupled MRF framework for simultaneous characterization and segmentation of the tumor. The approach is guided by the appropriate cluster's prior probability map. The method differs from existing atlas-based tumor segmentation methods since our atlas describes the expected position of the tumor (instead of detecting the tumor with respect to the healthy structures [Gering 2002, Prastawa 2004, Menze 2010]) and independent from the observed image (contrarily to the work of [Prastawa 2003] that modifies a healthy atlas based on the observation).

3.3.1 Tumor Detection

The first step towards tumor segmentation and characterization is an initial detection of the tumor to roughly estimate its position. We rely on boosting due to the simplicity of the Adaboost algorithm [Freund 1997], and its computational efficiency once the classifier is learned.

3.3.1.1 Adaboost Algorithm

Consider an ensemble of labeled volumes $\mathcal{X}$ featuring a brain tumor. N voxels are extracted from the different volumes to construct a training set that consists of pairs $\{\mathbf{x_i}, y_i\}, i \in [1:N]$ where $\mathbf{x_i} \in \mathcal{X}$ is an observed image voxel and y_i is its corresponding label (-1 for background, 1 for tumor). Each pair is associated to a feature vector $f(\mathbf{x_i})$ and a weight D_i. The idea of boosting

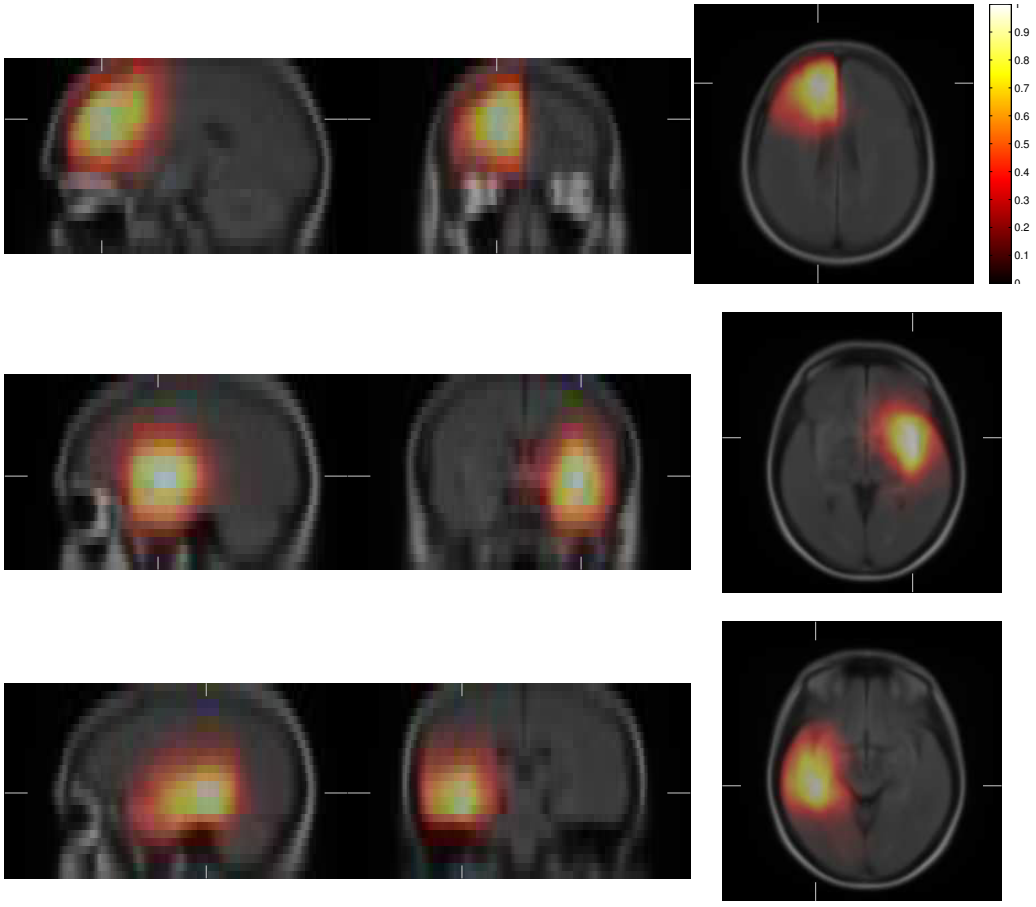


Figure 3.3: *Probability maps of three different clusters superimposed to the mean registered image.*

is to construct a strong classifier H as a linear combination weakly discriminative classifiers (the weak learners). It is an iterative process, where a new weak learner is constructed at each iteration. The weights D_i determine the importance of their corresponding sample assignment during classification. In each step of the process, a each weak learner is built, and the samples' weights are updated so that the focus is on the misclassified and more difficult samples at the next iteration.

Initially, all weights are set uniformly: $D_i^0 = 1/N$ while the user specified the number of iterations T and the kind of weak learner used. At each iteration t, a new weak learner $h_t : \mathcal{X} \rightarrow \{-1, 1\}$ is constructed based on a value in the feature vector in order to minimize the classification error

$$\epsilon_t = \sum_{i=1}^{N} D_i^t |h_t(\mathbf{x}_i) - y_i| \tag{3.17}$$

Once the classifier has been chosen, its importance with respect to the strong classifier is computed via the weighting parameter α_t and the weights are updated in order to enhance the focus on misclassified samples:

$$\alpha_t = \frac{1}{N}\log(\frac{1-\epsilon_t}{\epsilon_t}) \tag{3.18}$$

$$D_i^t = \frac{1}{Z_i} D_i^{t-1} \exp(-\alpha_t y_i h_t(\mathbf{x_i})) \tag{3.19}$$

where Z_i is a normalizing constant. Eventually, the strong classifier is computed as:

$$H(\mathbf{x}) = \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}) \tag{3.20}$$

Among nice properties of the AdaBoost algorithm, it was shown in [Freund 1997] that the classification error rate decreases exponentially fast while [Friedman 2000] showed that the boosting score could be converted to probabilities as:

$$\begin{aligned} P(y=1|\mathbf{x}) &= \frac{e^{H(\mathbf{x})}}{e^{-H(\mathbf{x})}+e^{H(\mathbf{x})}} \\ P(y=-1|\mathbf{x}) &= \frac{e^{-H(\mathbf{x})}}{e^{-H(\mathbf{x})}+e^{H(\mathbf{x})}} \end{aligned} \tag{3.21}$$

In [Friedman 2000], a variant of the AdaBoost algorithm called *Gentle Adaboost* was introduced. The idea was to provide a method that is less sensitive to outliers and noise. Indeed, the use of log-ratios in the evaluation of α_t can be numerically unstable and lead to an excessive focus on outliers. The Gentle Adaboost algorithm has been shown to yield more robust results and proceeds as follows:

1. Initialize weights $D_i = 1/N, i \in [1:N]$ and classifier $H(\mathbf{x}) = 0$
2. Repeat for $t = 1, ..., T$
   (a) Find a weak classifier that minimizes $\epsilon_t = \sum_{i=1}^{N} D_i (y_i - h_t(\mathbf{x_i}))^2$
   (b) Update $H(\mathbf{x}) \leftarrow H(\mathbf{x}) + h_t(\mathbf{x})$
   (c) Update $D_i \leftarrow \frac{1}{Z_i} D_i \exp(-y_i h_t(\mathbf{x_i}))$
3. Output classification result: $\text{sign}\left(\sum_{t=1}^{T} h_t(\mathbf{x})\right)$

In this thesis, the Gentle Adaboost algorithm is used while the weak learners are defined as decision stumps.

3.3.1.2 Feature Vector

One of the key elements for a good separation between the tumor and background classes is the design of the feature vector that has to be discriminative. We consider iconic, texture and symmetry information.

- Iconic features rely on the intensity values of the image and the fact that the tumor has a different appearance than the background. In order to integrate neighborhood information, $9 \times 9 \times 5$ intensity patches centered on the voxel of interest are added to the feature vector. The patches are constructed by selecting one voxel out of two in the image for increased neighborhood exploration. Rotation invariant statistics (median, standard deviation and entropy) on patches of size $k \times k \times 3$ where $k = [3, 5, 7]$ are also included. The smaller patch size in the z axis is due to the lower resolution in this direction.

- Texture information is added via the computation of Gabor filters [Manjunath 1996]. Gabor filters have been used in various medical image analysis problems in order to detect the texture patterns in the image. A 2-D Gabor function is a sinusoidal plane wave of a certain frequency λ and orientation θ (the carrier) modulated by a Gaussian envelope of scale $\sigma = (\sigma_x, \sigma_y)$:

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right) + i\frac{2\pi}{\lambda}(x' + y')\right) \quad \text{where } x' = x\cos\theta \text{ and } y' = y\cos\theta \tag{3.22}$$

  Convolution of the Gabor function to an image yields a high response when the local image structure is similar to the Gabor function based on its scale, wavelength and orientation. A set of Gabor functions $G_{\sigma,\theta_n}(x,y)$ (filter bank) can be constructed by selecting a number of scales S and orientations K so that it thoroughly paves the frequency space. In practice, 3D computation of the Gabor filters is very slow. We adopt the method proposed in [Zhan 2003] that approximates the 3D filter by computing two orthogonal 2D filter banks. We use the fast rotation and scale invariant implementation used in [Michel 2011].

  It was observed that texture features have little discriminative power with respect to the tumor. However, they are very efficient towards determining the brain's main structures and consequently reducing the risk of false detections. The filters are computed on 2 scales and 10 orientations. The image response to each obtained filter at voxel $\mathbf{x}$ constitutes one element of the feature vector $f(\mathbf{x})$.

- The last feature used is specific to the brain and exploits the approximative symmetry between the two hemispheres. The idea is that the presence of the tumor will introduce a notion of dissymmetry. Assuming that an approximate symmetry plane Π is known, the symmetry measure evaluates the difference between the intensity of a voxel $\mathbf{x}$ in the left part of the image and its symmetric $\mathbf{x_\Pi}$. To compensate the approximation in the selection of

the symmetry plane, the mean intensity of a patch around $\mathbf{x}$ and $\mathbf{x_\Pi}$ is computed instead of using the single intensity value. The symmetry feature is thus computed as follows:

$$\begin{aligned} S(\mathbf{x}) &= \frac{1}{N}\sum_{\mathbf{y}\in\mathcal{P}(\mathbf{x})} I(\mathbf{y}) - \frac{1}{N}\sum_{\mathbf{y_\Pi}\in\mathcal{P}(\mathbf{x_\Pi})} I(\mathbf{y_\Pi}) \\ S(\mathbf{x_\Pi}) &= -S(\mathbf{x}) \end{aligned} \tag{3.23}$$

where $\mathcal{P}(\mathbf{x})$ corresponds to the patch centered around $\mathbf{x}$ and N is the number of voxels in $\mathcal{P}(\mathbf{x})$. Given that the volumes to be segmented are registered to the same reference space where the atlas was built, estimating an approximate symmetry plane is straighforward.

Visual examples of the different elements constituting the feature vector can be seen in Fig. [3.4] while examples of boosting classification results are shown in Fig.[3.15a]

3.3.2 Coupled Markov Random Field Model

We will now proceed with the presentation of the Markov Random Field model that couples tumor characterization and segmentation with spatial priors. We formulate the problem of tumor segmentation and characterization as a labeling problem where we want to assign a global volume wise characterization label $c(\mathbf{x})$ and a voxel wise segmentation label $s(\mathbf{x})$ to each image voxel. The characterization label corresponds to assigning the patient/tumor to one of the atlas' K clusters $c(\mathbf{x}) \in C_1, ..., C_K$, *i.e.* determining which preferential location the tumor is situated in. Given the assumption that the patient has only one tumor, the same label has to be assigned to all the voxels. The segmentation label $s(\mathbf{x}) \in \{0, 1\}$ simply characterizes the voxel as being part of the tumor ($s(\mathbf{x}) = 1$) or the background ($s(\mathbf{x}) = 0$).

We adopt a pairwise MRF model on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where each node $\mathbf{x} \in \mathcal{V}$ corresponds to an image voxel while the edges constitute the neighborhood $\mathcal{N}(\mathbf{x})$ by connecting node $\mathbf{x}$ to its 6 immediate neighbors in 3D. A node $\mathbf{x}$ is to be assigned a label $l_x \in \mathcal{L} = \{1, \cdots, K\} \times \{0, 1\}$ that jointly characterizes and segments the tumor. We map this label space into one dimension as follows:

$$\begin{aligned} c(\mathbf{x}) &= c^{l_x} = C_k \text{ where } k = l_x \pmod K \\ s(\mathbf{x}) &= s^{l_x} = \left\lceil \frac{l_x}{K} \right\rceil \end{aligned} \tag{3.24}$$

The optimal labeling $\mathbf{l} = \{l_x | \mathbf{x} \in \mathcal{V}\}$ is recovered by minimizing the MRF energy defined as:

$$E(\mathbf{l}) = \sum_{\mathbf{x}\in\mathcal{V}} V_x(l_x) + \sum_{\mathbf{x}\in\mathcal{V}}\sum_{\mathbf{y}\in\mathcal{N}(\mathbf{x})} V_{xy}(l_x, l_y) \tag{3.25}$$

3.3.2.1 Pairwise Term

The pairwise term acts as a smoothing prior. Its role is two-fold: it imposes local spatial consistency of the voxel wise segmentation and ensures that the same characterization label is assigned to the

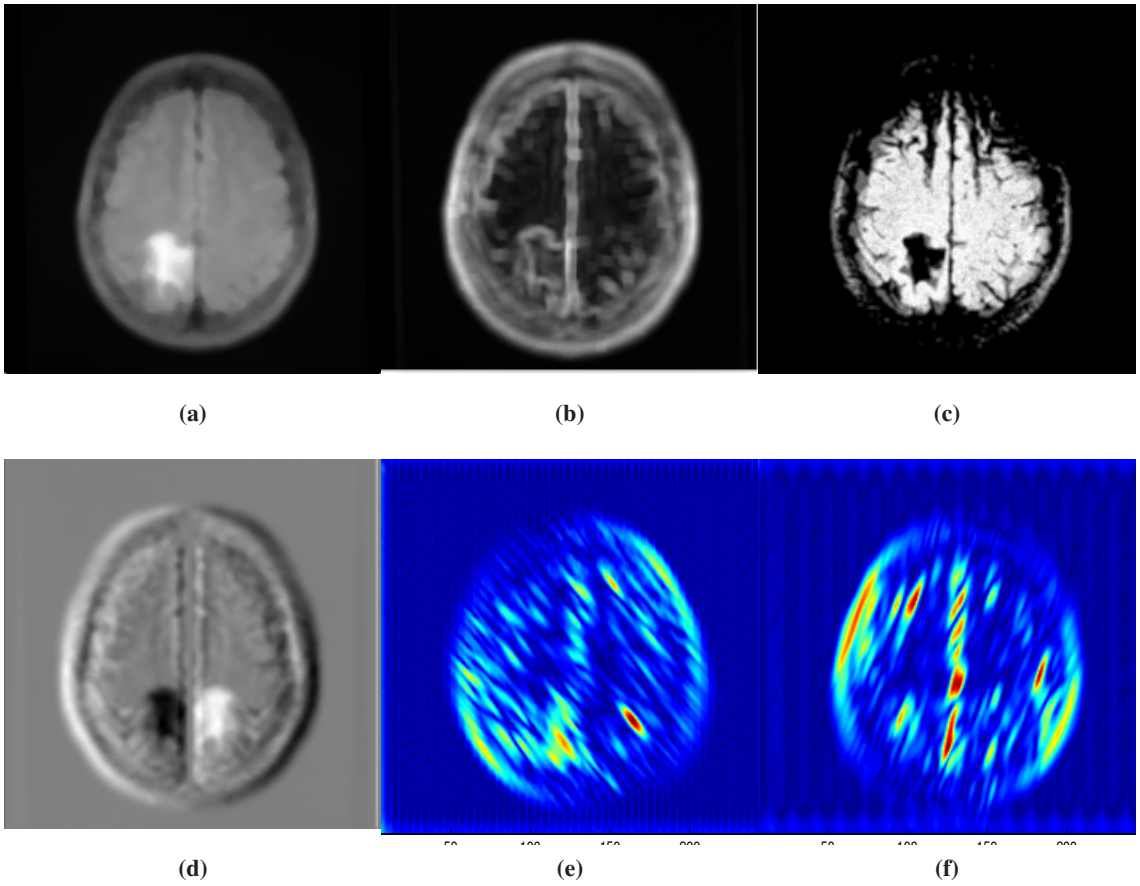


Figure 3.4: *Visual examples of features used for boosting training: (a) median filter, (b) standard deviation, (c) entropy, (d) symmetry, (e,f) Gabor features.*

whole volume. It is designed following the Potts model, by penalizing neighboring nodes that have been assigned different labels:

$$V_{xy}(l_x, l_y) = \begin{cases} 0, & \text{if } l_x = l_y \\ \beta, & \text{if } s^{l_x} \neq s^{l_y} \text{ and } c^{l_x} = c^{l_y} \\ \gamma, & \text{otherwise} \end{cases} \tag{3.26}$$

Where β is a constant parameter that describes the strength of the penalization imposed on the segmentation field and $\beta \ll \gamma$.

3.3.2.2 Unary Terms

We define the MRF unary term as a sum of three potentials :

$$V_x(l_x) = V_S(s^{l_x}) + \gamma V_C(c^{l_x}) + \mu V_{S,C}(s^{l_x}, c^{l_x}) \tag{3.27}$$

The first two terms V_S and V_C are observation dependent and focus on the segmentation and characterization space respectively, the third term $V_{S,C}$ couples the two, while γ and μ are constant parameters that balance the relative importance of the different potentials.

For the segmentation term $V_S(s^{l_x})$, we make use of the classification likelihoods $P_{tm}(I(\mathbf{x}))$ and $P_{bg}(I(\mathbf{x}))$ recovered through Gentle AdaBoost learning. This terms seeks the label configuration that yields the maximum likelihood probability. Inspired by the MAP-MRF formulation and following the approach of [Boykov 2006], we use the log probabilities and define the term as follows:

$$V_S(s^{l_x}) = -s^{l_x} \log P_{tm}(I(\mathbf{x})) - (1 - s^{l_x}) \log P_{bg}(I(\mathbf{x})) \tag{3.28}$$

The characterization term's role is to identify which cluster the tumor to be segmented belongs to. Even though the precise position of the tumor is unknown, the boosting classification score $H(\mathbf{x})$ gives a rough estimate of the position of the tumor pixels. The closest cluster is identify by measuring the amount of non overlap between the binary map associated to the center of each cluster and the voxels with a high boosting classification score.

$$V_C(c^{l_x}) = \max(0, H_T(I(\mathbf{x})) - \bar{M}_{c^{l_x}}(\mathbf{x})) \tag{3.29}$$

Where $\bar{M}_{C_i}(.)$ is the binary segmentation map of the tumor that is the center of the cluster C_i, and $H_T(.)$ is a binary map representing voxels with a high classification score, obtained by thresholding (with threshold T) of the boosting classifier output:

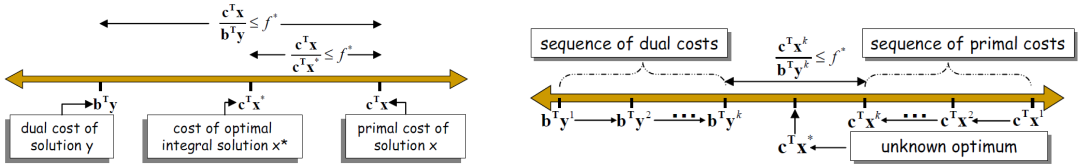
$$H_T(I(\mathbf{x})) \begin{cases} 1, \text{ if } H(I(\mathbf{x})) > T \\ 0, \text{ otherwise} \end{cases} \tag{3.30}$$

This potential counts the number of voxels that are classified as tumor by boosting and do not overlap with the tumor volume of the cluster center. The closest cluster corresponds to the one that has the highest overlap with the most probable tumor voxels.

Last but not least, the coupling term $V_{S,C}(s^{l_x}, c^{l_x})$ links the segmentation and characterization labels via the DLGG atlas and plays an essential role in the segmentation process. Let us recall that each cluster C_i is associated to a probability map $P(\mathbf{x}|C_i)$ that describes the spatial distribution of tumor voxels in the cluster. Given a characterization label (cluster assignment), the corresponding probability map gives insight on where the tumor is expected to be and not to be. This is used as a spatial prior that will encourage or discourage segmentation decisions based on the voxel's localization:

$$V_{S,C}(s^{l_x}, c^{l_x}) = -s^{l_x} \log P(\mathbf{x}|c^{l_x}) - (1 - s^{l_x}) \log \left(1 - P(\mathbf{x}|c^{l_x})\right) \tag{3.31}$$

Spatial consistency is enforced between the segmentation label and the expected spatial distribution of the tumors of the corresponding cluster.



(a) (b)

Figure 3.5: *Image courtesy of [Komodakis 2008b]. (a) Primal-Dual principle: the optimal cost always lies between* $\mathbf{b}^T\mathbf{y}$ *and* $\mathbf{c}^T\mathbf{x}$ *for any pair* $(\mathbf{x}, \mathbf{y})$ *of feasible solutions. If both values are close, the approximated solution to the primal problem is close to the optimal solution too. (b) Primal-Dual schema: iteratively generate pairs* $(\mathbf{x}^k, \mathbf{y}^k)$ *of feasible solution until they are close enough (their ratio is below* $f*$*) to obtain an* $f*$*-approximation of the solution*

3.3.2.3 MRF Optimization Using Linear Programming

The MRF energy is minimized using Fast-PD [Komodakis 2007b, Komodakis 2008b]. The algorithm relies on linear programming and primal-dual strategies. Considering a pair of primal and dual linear programs:

$$\begin{aligned} \text{PRIMAL} : \min \mathbf{c}^T\mathbf{x} \qquad & \qquad \text{DUAL} : \quad \max \mathbf{b}^T\mathbf{y} \\ \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \qquad & \qquad \text{s.t. } \mathbf{A}^T\mathbf{y} \leq \mathbf{c} \end{aligned} \tag{3.32}$$

It is possible to obtain an f-approximation of the solution to the Primal problem by minimizing the cost gap between the primal and dual problems so that their ratio is inferior to f. This is carried out by iteratively generating pairs of primal-dual solutions $\{\mathbf{x}^t, \mathbf{y}^t\}$ until the elements of the last generated pair are both feasible and have a primal dual gap inferior to f:

$$\mathbf{c}^T\mathbf{x}^t \leq f.\mathbf{b}^T\mathbf{y}^t \tag{3.33}$$

This strategy is applied to the MRF energy by reformulating the problem as the following integer program:

$$\min \sum_{p\in\mathcal{V}} \sum_{l\in\mathcal{L}} V_p(l)x_p(l) + \sum_{p,q\in\mathcal{E}} \sum_{l,l'\in\mathcal{L}} V_{pq}(l,l')x_{pq}(l,l') \tag{3.34}$$

$$\text{s.t. } \sum_l x_p(l) = 1 \qquad \forall \mathbf{p} \in \mathcal{V} \tag{3.35}$$

$$\sum_l x_{pq}(l,l') = x_q(l') \qquad \forall l' \in \mathcal{L}, (\mathbf{p},\mathbf{q}) \in \mathcal{E} \tag{3.36}$$

$$\sum_{l'} x_{pq}(l,l') = x_p(l) \qquad \forall l \in \mathcal{L}, (\mathbf{p},\mathbf{q}) \in \mathcal{E} \tag{3.37}$$

$$x_{pq}, x_p \in \{0,1\} \tag{3.38}$$

The binary variables $x_p(l)$ and $x_{pq}(l,l')$ characterize the label assignments. $x_p(l) = 1$ implies that label l has been assigned to node $\mathbf{p}$ while $x_{pq}(l,l') = 1$ indicates that nodes $\mathbf{p}$ and $\mathbf{q}$ are assigned label l and l' respectively. Constraint 3.35 imposes that the node is only assigned one label and constraints 3.36 and 3.37 impose consistency between the unary and pairwise binary variables. The linear programming relaxation of the integer program consists in relaxing constraint 3.38 to the constraints $x_p(.) \geq 0$ and $x_{pq}(.,.) \geq 0$. The primal-dual scheme is then applied to the relaxed integer program and its corresponding Dual problem:

$$\max \sum_{p \in \mathcal{V}} y_p \tag{3.39}$$

$$\text{s.t. } y_p \leq \min_{l \in \mathcal{L}} h_p(l) \qquad \forall p \in \mathcal{V} \tag{3.40}$$

$$y_{pq}(l) + y_{qp}(l') \leq V_{pq}(l,l' \qquad \forall (l,l') \in \mathcal{L}, (\mathbf{p},\mathbf{q}) \in \mathcal{E} \tag{3.41}$$

$$\tag{3.42}$$

where $h_p(.) = V_p(.) + \sum_{q,(p,q)\in\mathcal{E}} y_{p,q}(.)$ is called the height variable. For MRF optimization, each iteration (generation of primal-dual pairs) corresponds to a max-flow optimization of a capacitated graph constructed based on the current pair of primal-dual variables.

The method has several advantages among which its generality (the pairwise potentials only have to be positive), optimality guarantee (f-approximation of the optimum) and computational efficiency.

3.4 Experimental Validation

3.4.1 Data-set and Preprocessing

The database used to build the statistical atlas consists of 210 carefully selected patients. All of them were over 18 years old and treated in the same institution. A histological diagnosis has been performed and FLAIR volumes as well as molecular analysis where available prior to any treatment. The tumor size was limited to $180cm^3$ so that the center of the tumor remains representative of where the tumor initially appeared: the tumor expansion would shift its center of gravity. The distribution of tumors' sizes is shown in Fig. [3.6]. The database consisted of 97 females and 113 males of mean age 38 years old and mean tumor volume 45 cm^3. The tumor size ranged from 0.1 to 180 cm^3.

The patients' MRIs considered for the construction of the atlas were 3D FLAIR volumes that were all manually segmented. The image size varied from 144 to 576 in the (x,y) plane et 12 to 213 in the z plane, and the voxel resolution from 0.4x0.4 to 1x1 mm^2 in the (x,y) plane and 0.9 to 10 mm in the z plane.

The reference pose used for registration is a 256x256x24 FLAIR image of a healthy brain, with resolution 0.9x0.9x5.5. Because of the strong slice thickness, the segmentation maps were no longer binary after registration due to interpolation tasks. The same threshold of 0.2 determined

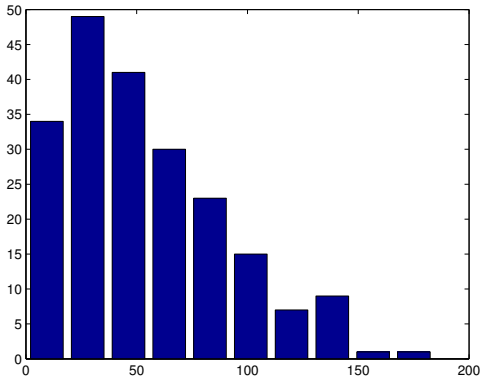


Figure 3.6: *Distribution of tumors sizes in the data set (number of tumors vs tumor size in cm^3)*

experimentally as the most faithful segmentation of the registered tumor is used on all the data and in the remainder of this thesis to convert the registered maps to binary maps.

The boosting features rely heavily on the images intensities that need to be comparable between volumes. The intervolume intensity inhomogeneity inherent to the MRI modality is dealt with using a simple method, in order to maintain the contrast between the tumor and the background. All images are set to a similar intensity range by setting them to the same median and interquartile range:

$$I_{reg} = \frac{I - med(I)}{IQR(I)} \tag{3.43}$$

where med(I) and $IQR(I)$ are respectively the median and interquartile range of image I, computed without taking into account the background area (lowest intensity values). The median was used instead of the mean value due to the presence of the tumor that can impact the mean intensity value, especially if the tumor's size is significant. No correction was done on the image's bias field.

3.4.2 Sparse Graph Validation and Robustness

In order to select the best parameter α for clustering, we clustered the complete graph for different values of α ranging from 0.3 to 7 which yields 5 to 49 clusters. For each experiment, we computed the value of the three indices (Dunn, Davies-Bouldin and Silhouette). All indices showed a peak corresponding to 11 clusters. The indices have been normalized and combined using their product: $I_c = (D.(1 - DB).GS)^{1/3}$ with the underlying idea that if an index has a low value, the combined value should also be low. The combined index value showed a strong maximum for $\alpha = 2.1$ which corresponds to a 11 clusters graph. The different indices values as well as their combined value with respect to parameter α and the number of clusters can be seen in Fig. [3.10]. The clustered graph is shown in Fig. [3.7]. The positions of the centers of the clusters were transferred to the ICBM atlas via non rigid registration of the reference pose [Glocker 2008b] and can be seen in

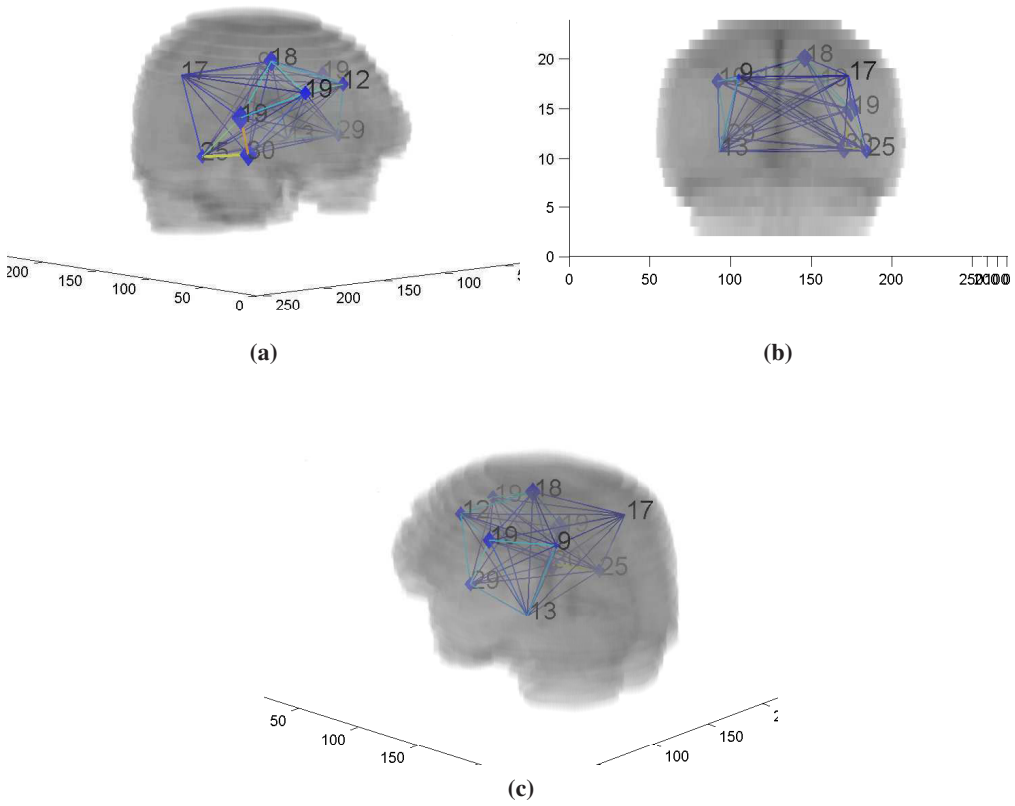


Figure 3.7: *Visualization of the complete clustered graph superimposed to the mean registered image. The numbers correspond to the number of nodes in each cluster.*

Fig. [3.14], highlighting a strong symmetry between the hemispheres. We observe, as summarized in table 3.1, a strong predominance of tumor appearances in the Insula (more than a third of the tumors) and towards the frontal lobe, while there are no clusters in the occipital lobe. Schematics of the brain structure are shown in Fig. [3.8]

Next, we evaluated the robustness of the graph by performing Leave One Out Cross Validation. We carried out 210 experiments where one sample $\mathbf{p}$ was removed from the data set. For each experiment, we learned the graph topology on the 209 remaining samples by selecting the clustering that yielded the maximum value of the three indices. An 11 clusters graph was obtained for 207 experiments (98.5 % of the data). A 10 clusters graph was obtained for 3 samples. Those samples were the center or close to the center of the same loose cluster that was merged with its closest neighbor when the sample was removed.

We then studied the robustness of the graph by analyzing the influence of each sample. For

Anatomical localization	Number of clusters	Percentage of data
Parietal lobe	2 (symmetric)	12 %
Supplementary Motor Area (frontal lobe)	1	9 %
Temporal lobe	2 (symmetric)	18 %
Frontal lobe	1	9 %
Anterior frontal (frontal lobe)	2 (symmetric)	15 %
Insula	3 (2 symmetric clusters)	37 %

Table 3.1: *Anatomical localization of the different clusters*

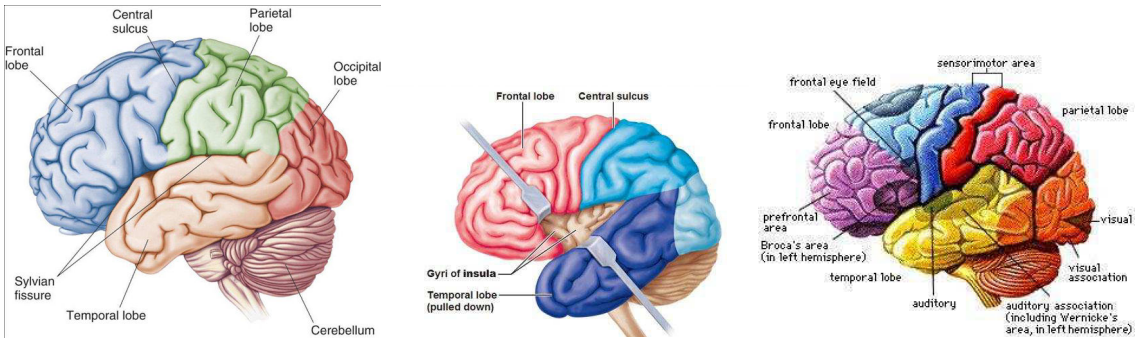


Figure 3.8: *Schematics of the brain structure.*

each learned clustered graph obtained by cross validation, we evaluated the impact of the 210th sample on the structure of the graph and the quality of the clusters. In order to determine how well the graph represents the data, we tried to assign the 210th sample to a cluster based on two measures: the distance from the center of the cluster and the silhouette index. To assign a sample to a cluster, three criteria have to be verified:

1. *Distance from the center of the cluster.* The sample has to be assigned to its closest cluster and its distance to the center should follow the distance distribution within the cluster. We determine if the sample is an outlier based on the distribution's quartiles :

$$d(\mathbf{p}, \bar{c_k}) = \min_{j \in [1:K]} d(\mathbf{p}, \bar{c_j}) \text{ and } d(\mathbf{p}, \bar{c_k}) \leq Q_3(C_k) + 1.5(IQR(C_k)) \tag{3.44}$$

   where $Q_3(C_k)$ and $IQR(C_k)$ are respectively the third quartile and interquartile range of the distance distribution in cluster C_k.

2. *Individual Silhouette index.* The global silhouette index is indicative of the quality of the clustering. The individual index tells how well the sample fits in its cluster. For each cluster k, we compute the corresponding silhouette index. If the index is close to 1, there is no doubt about the fact that the sample belongs to this cluster. A value close to zero does not

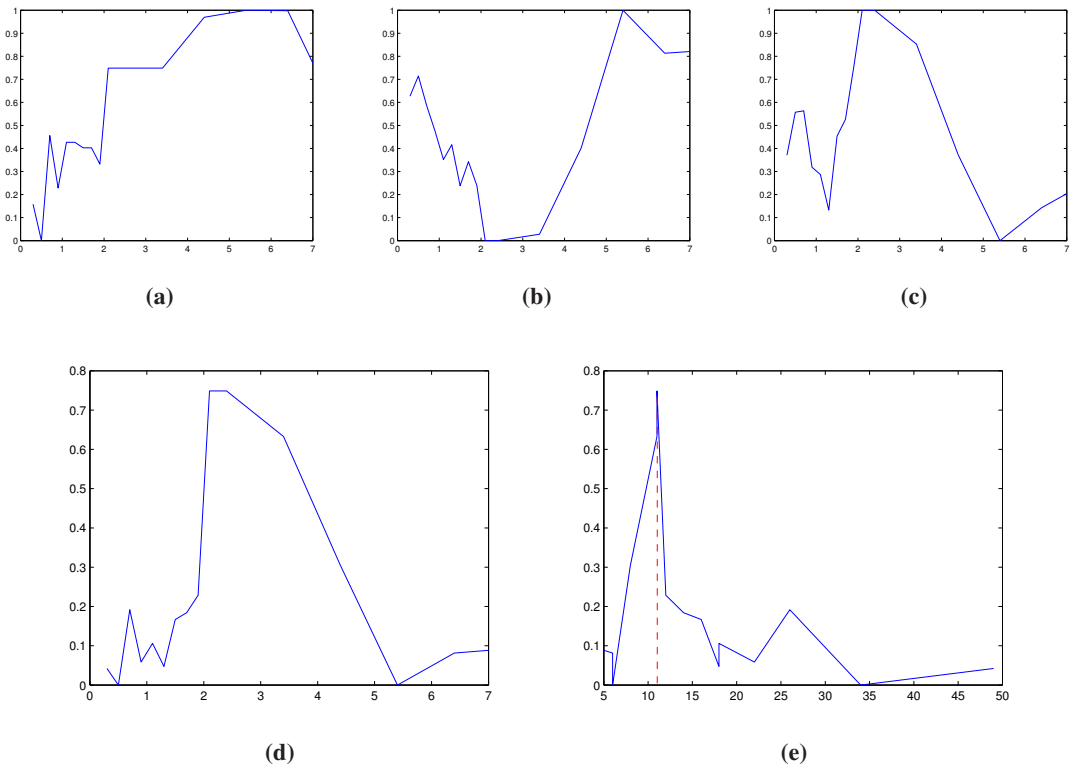


Figure 3.9: *Cluster validity indices with respect to the value of α (a,b,c,d) and the number of clusters (e). Dunn index (a), Davies Bouldin index (b), Silhouette index (c) and combined indices (d,e).*

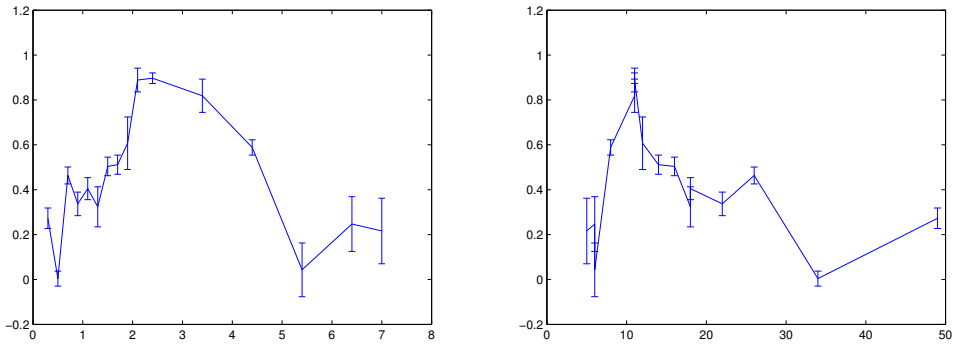
necessarily mean that the clustering is bad. Indeed, the main drawback of the silhouette is that it can penalize loose clusters (large diameter) when compared to smaller ones. The sample should be assigned to a cluster if it maximizes the silhouette index and fits in the silhouette of the cluster. Similarly to the first criterion, we make use of the quartile values:

$$s_k(\mathbf{p}) = \max_{j \in [1:K]} s_j(\mathbf{p}) \text{ and } s_k(\mathbf{p}) \geq Q_1(C_k) - 1.5(IQR(C_k)) \tag{3.45}$$

where $Q_1(C_k)$ and $IQR(C_k)$ are respectively the first quartile and interquartile range of the silhouette index distribution in cluster C_k and $s_k(\mathbf{p})$ is the silhouette index of sample $\mathbf{p}$ with respect to cluster C_k.

3. The last criterion verifies that the same cluster is selected by criteria 1 and 2:

$$\operatorname*{argmin}_k d(\mathbf{p}, \bar{c_k}) = \operatorname*{argmax}_k s_k(\mathbf{p}) \tag{3.46}$$



(a) **(b)**

Figure 3.10: *Errorbars of the indices computed for the cross validation experiments with respect to the value of α (a) and the number of clusters (b).*

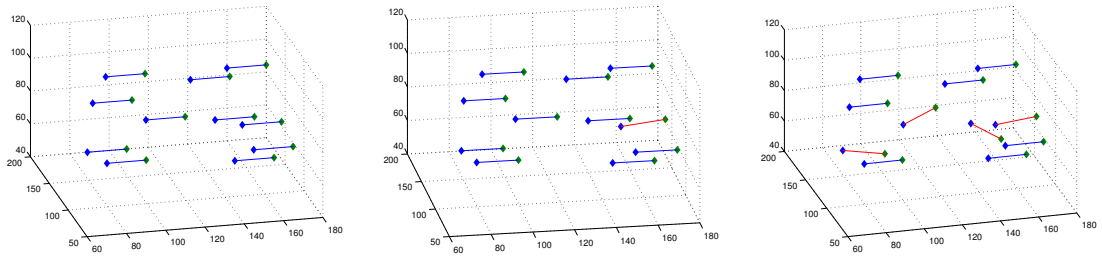
Using those criteria, we successfully classify 88% of the whole data-set.

Last, we evaluated topology correspondences between the 207 eleven-components clusters graphs and the graph obtained from the whole database. To this end, we use the graph matching algorithm presented in [Torresani 2008]. Complete match of the graphs was obtained in 70% of the cases and only one node differed in 26% of the cases. The rest of the samples had 2 (9 samples) or 4 (1 sample) different nodes. For the cases where the cluster centers were different, they were still situated in the same area in the brain and the centers were close. The maximum distance between two different centers is 12.6 which corresponds to the mean intra cluster distance of tight clusters. Complete and partial matching are illustrated in Fig. [3.11] that highlights the proximity of mismatched centers. The worst case with 4 non matched samples is shown in Fig. [3.11c].

3.4.3 Segmentation Results

Segmentation was performed on the 210 volumes via Leave One Out Cross Validation. The atlas was constructed on the 209 remaining samples according to the optimal value of α. For boosting learning, 210 classifiers were learned by randomly selecting 35 volumes among the 209 samples used for graph construction. For the symmetry feature, we rely on the fact that all images are registered and use an approximate symmetry plane estimated on the reference pose. Estimating an exact symmetry plane is out of the scope of this work and has been the subject of elaborated methods [Prima 2002].

We evaluated the segmentation quality with respect to the manual segmentations by computing the Dice Score (D), False (FP) and True positive (TP) rates and Mean Absolute Distance between contours (MAD). Given a manual segmentation M and an automatic segmentation A, the different



(a) (b) (c)

Figure 3.11: *Illustration of the graph matching results. (a) Complete match, (b,c) Partial match. Positive matches correspond to blue edges and mismatched samples to red edges.*

scores are defined as follows:

$$\begin{aligned} D &= \frac{2\|A \cap M\|}{\|A \cup M\| + \|A \cap M\|} \\ TP &= \frac{\|A \cap M\|}{\|M\|} \\ FP &= \frac{\|A\| - \|A \cap M\|}{\|A\|} \\ MAD &= \frac{1}{2}\left(\frac{1}{n_A}\sum_{x \in A} d_{min}(x, M) + \frac{1}{n_M}\sum_{x \in M} d_{min}(A, x)\right) \end{aligned} \tag{3.47}$$

where n_A and n_M are the number of voxels in A and M respectively and $d_{min}(x, M) = \inf(d(x, y)|y \in M)$ is the smallest distance of a point x to any point in M. The distance $d(.,.)$ used here is the euclidean distance. Basically, the MAD score characterizes how close the two contours are.

The method was compared to sole boosting classification and pairwise MAP-MRF segmentation using the boosting probabilities as likelihood and an Ising model as pairwise cost. The parameter β determining the amount of smoothing of the segmentation field was set to 1 with prior and 3 without prior (MAP-MRF). Parameters γ and μ were set to 1 and 10000 respectively. The high value of μ was necessary for small tumors that caused very little variation in the value of the potential $V_C(.)$. The threshold parameter T driving the construction of the binary boosting score is set to 1.5. The parameters μ and T were determined so that the maximum number of images were assigned to their right cluster (inferred from the complete graph construction).

Using our framework, we observed an increase in the quality of the segmentation and a significant drop of false detections. The different mean and median scores for the three methods are shown in table 3.2 and boxplots of the scores over all images are shown in Fig. [3.12]. We observe an increase of the Dice score (+ 3% mean and median) and strong reduction of the false positive rate (- 9-10 % mean and median). The increase in quality is highlighted by the MAD score, as illus-

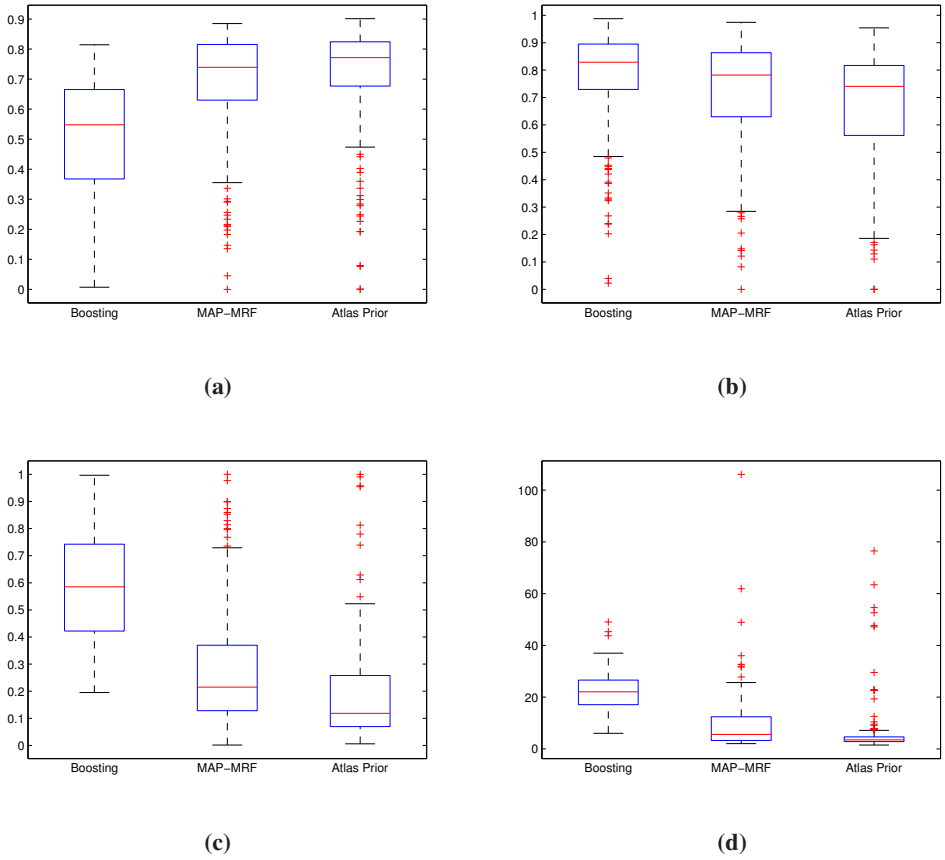


Figure 3.12: *Boxplots of the Dice score (a), True Positive rate (b), False positive rate (c), and MAD score (d) between the automatic and manual tumor segmentation for the three different methods.*

trated in the boxplot in Fig. [3.12d]. Visual segmentation results shown in Fig.[3.15] also highlight the impact of the spatial prior, where erroneous boosting detections are corrected. A reduction of the true positive rate is observed (-4 %), mainly due to mischaracterized tumors and outliers as well as the limited number of samples in certain clusters. This phenomenon could be reduced by an increase in the size of the data set and detecting outlier tumors (as tumors equally distant from two cluster for instance) that should not be assigned to any cluster. Wrong clusters assignments are directly connected to the quality of the detection step and principally concern small tumors. Increasing the characterization quality would mostly depend on an increase of the quality of the detection. The discrepancy between the mean and median Dice values (69% and 77%) was due to those small tumors, that were poorly detected by the classifier and therefore poorly segmented.

Method	Dice	True Positive	False Positive	MAD
Boosting	50 (55) %	77 (83) %	59 (58) %	22 (22) mm
MAP-MRF	66 (74) %	70 (78) %	29 (21) %	9.5 (6) mm
Atlas Prior	69 (77) %	66 (74) %	20 (12) %	6 (3.5) mm

Table 3.2: *Mean and median (in parentheses) values of the different scores and methods for tumor segmentation.*

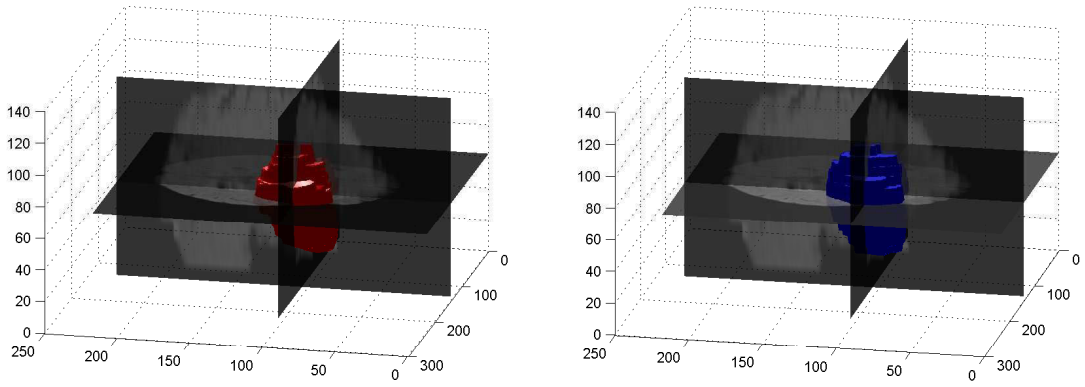


Figure 3.13: *3D visualization of an example segmentation result. (a) manual segmentation, (b) automatic segmentation*

3.5 Discussion

In this chapter, the construction of a probabilistic atlas mapping the preferential locations of DLGGs has been presented. The atlas was constructed through a complete graph where each node represents a tumor and the edges represent the proximity between the tumors, evaluated as the Euclidean distance between their center of gravity. The DLGGs' preferential locations were then identified by regrouping nearby tumors in clusters where each cluster corresponds to a preferential location. The optimal clustering was determined using well-established cluster validity indices. To each cluster is associated a probability map representing the spatial extension of tumors within the cluster.

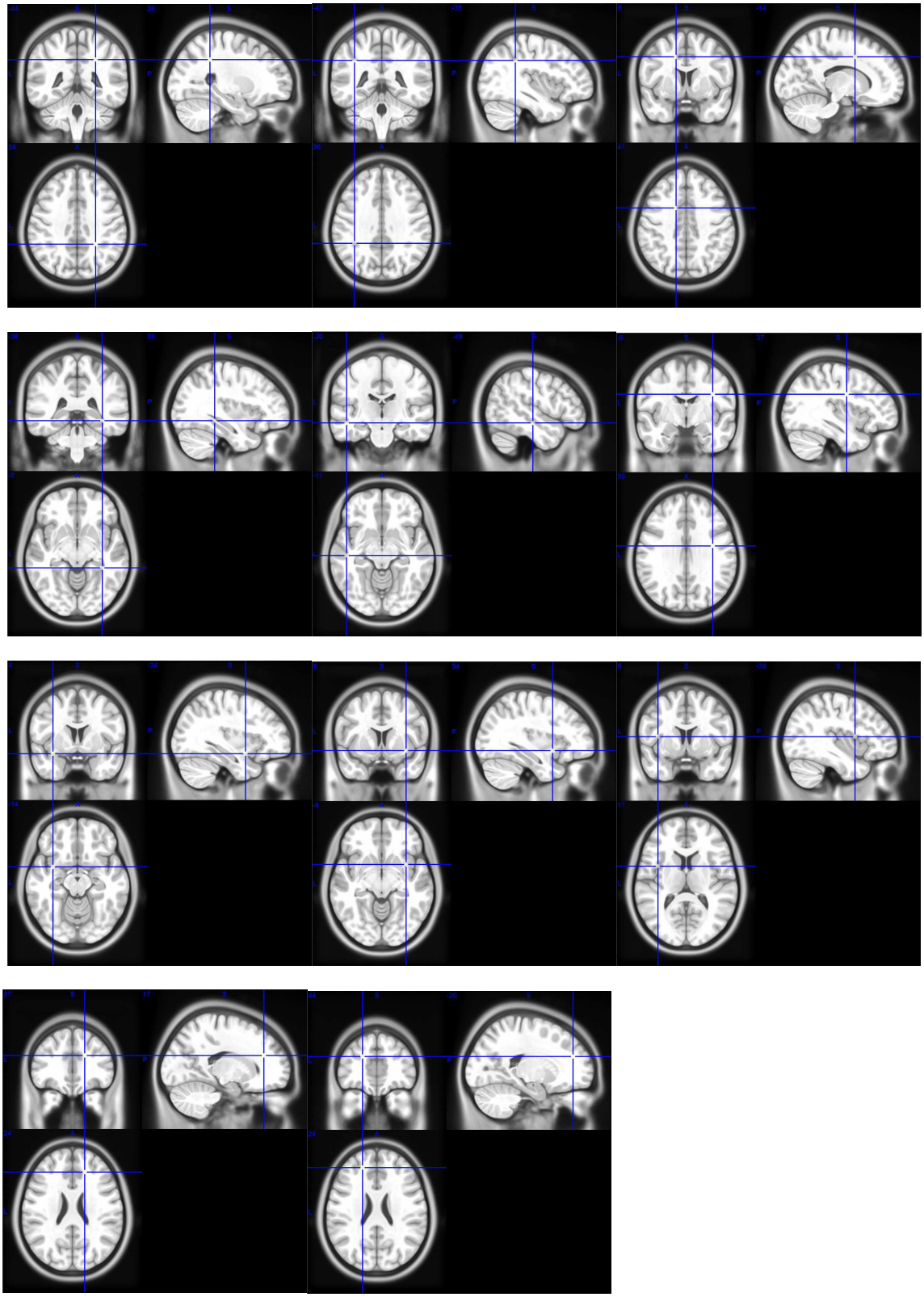
A natural application of this tumor specific atlas was in the context of tumor segmentation as a spatial prior. Such an application was presented in the second part of this chapter. The atlas was integrated in a coupled MRF framework where the tumor is concurrently characterized (assigned to one of the atlas' clusters) and segmented. An initial statistical classification based detection enables assignment to the cluster as well as segmentation with spatial smoothing (similarly to MAP-MRF approaches). The two problems (characterization and segmentation) are coupled via

a spatial prior term dependent on the characterization and segmentation labels. This term makes use of the cluster's probability maps to enforce spatial consistency with the segmentation and the expected localization of tumor pixels within a given cluster.

Aside from the obvious appeal of tumor segmentation, the atlas enables the study of the tumor behavior with respect to its localization which could be correlated with the patient's characteristics (age, sex...) or the tumor's characteristics (histological type, molecular biology...). This could potentially unravel behaviors (growth rate, anaplastic transformation) dependent on specific parameters and enable to adapt treatments to the patient. Going further, the interaction between the tumors and their surrounding tissue and functional areas as well as their growth patterns could be studied with respect to the cluster they're in. This could lead to models of tumor growth and their impact on brain functional reorganization.

The latter applications are however limited by the use of affine registration. Consequently, the interaction between the tumors and surrounding structures would strongly lack precision. A deformable registration could handle these issues, where the presence of the tumor has to be taken into account in order to obtain realistic transformations. Integration of anatomical information would also increase the quality of segmentation results that, in our approach, is strongly dependent on the adequacy of the selected cluster. An outlier tumor or a poorly detected tumor can be assigned to a cluster that does not correspond to its location. Anatomical information would reduce the impact of bad assignments as it offers a second means of integrating spatial prior knowledge.

The development of such a deformable registration method, coupled with tumor segmentation, will be the subject of the next chapter.



(a) (b)

Figure 3.14: *Positions of the cluster centers on the ICBM atlas after registration. A strong symmetry can be observed between clusters in columns (a) and (b) of the same image row.*

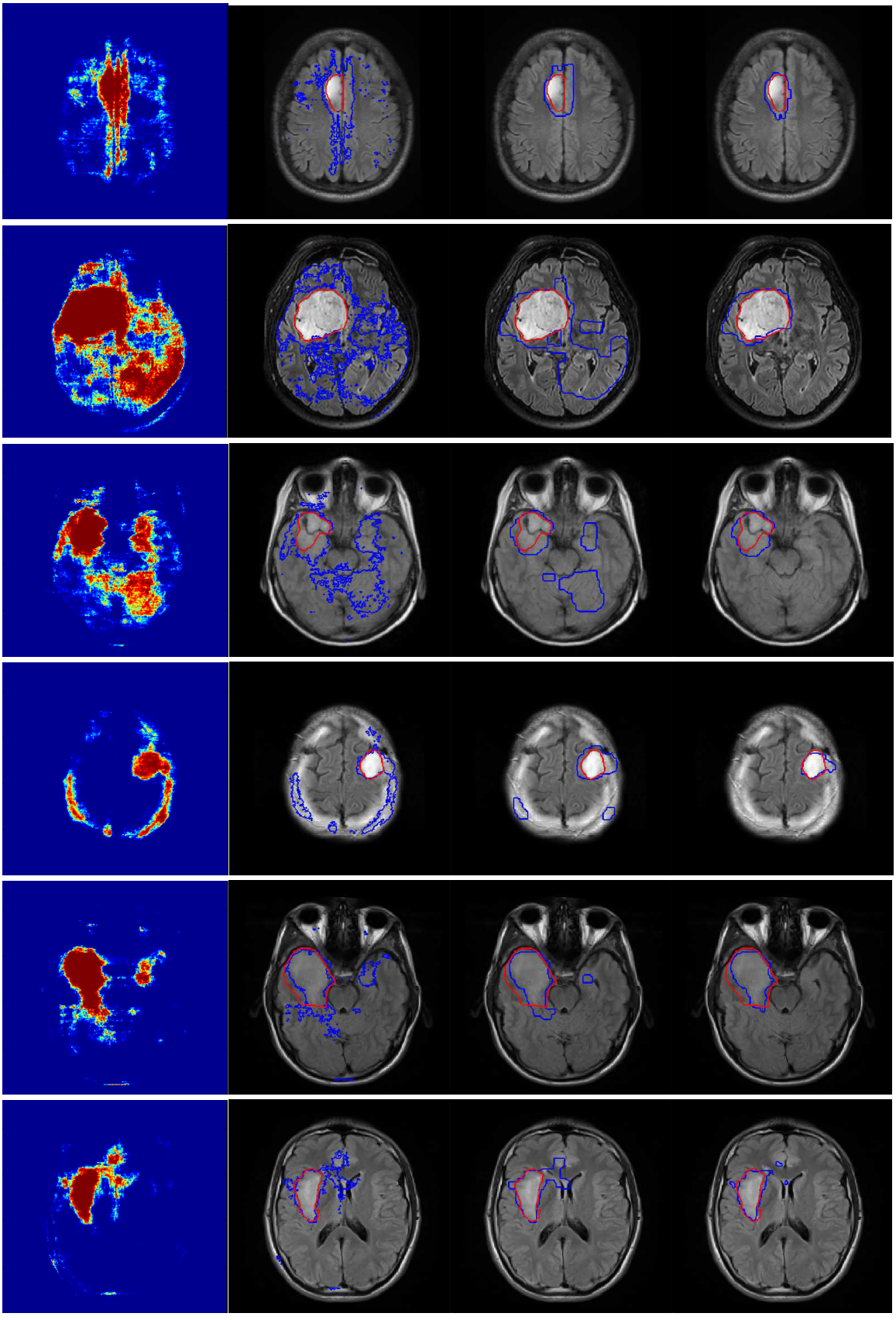


Figure 3.15: *Visual Segmentation results. (a) boosting score, (b) Boosting classification (thresholding), (c) Pairwise MRF, (d) MRF with spatial prior.*

Chapter 4

Joint Tumor Segmentation and Dense Deformable Registration

4.1 Introduction

In the previous chapter, a DLGG specific atlas was constructed in order to study the tumor appearances in the brain and their behavior. The methodology was designed to be robust to errors in registration, relying on the tumors center of gravity instead of surfaces. In general, the presence of tumors violates common registration procedures and therefore, a global affine registration was used to map all images in the same reference coordinate system. Interesting and promising results where obtained with respect to the mapping of DLGGs in the brain and their preferential locations. Nevertheless, an improvement on the precision is expected to be obtained via the use of deformable registration that would also provide the possibility of associating the presence of the tumor with the surrounding structures and in particular measure the impact of tumors on them. Generally speaking, when constructing pathology related atlases, the registration method should take into account the presence of the pathology and the absence of correspondence in the atlas.

In parallel to atlas construction, the task of atlas based segmentation of tumor bearing images has gained a lot of interest. A healthy reference atlas is registered to a pathological subject in order to segment its healthy structures and/or the tumor. For tumor segmentation (see chapter 2), a simple registration scheme is usually proposed (affine registration [Prastawa 2004], or segmentation based registration with the tumor labeled healthy [Kaus 2001]) and subsequently, a segmentation method is designed in order to cope with registration errors. However, such methods fail when the registration error is significant (i.e. when the brain is deformed by the tumor). If the healthy structures are to be segmented based on the atlas' structures segmentation, affine registration cannot match the different anatomies. Those issues highlight the need for a non-rigid registration scheme for tumor bearing images in order to obtain robust and precise segmentations.

Most of the existing methods for registration with missing correspondences consider that the tumor has been determined by a prior segmentation. The coupled segmentation/registration task

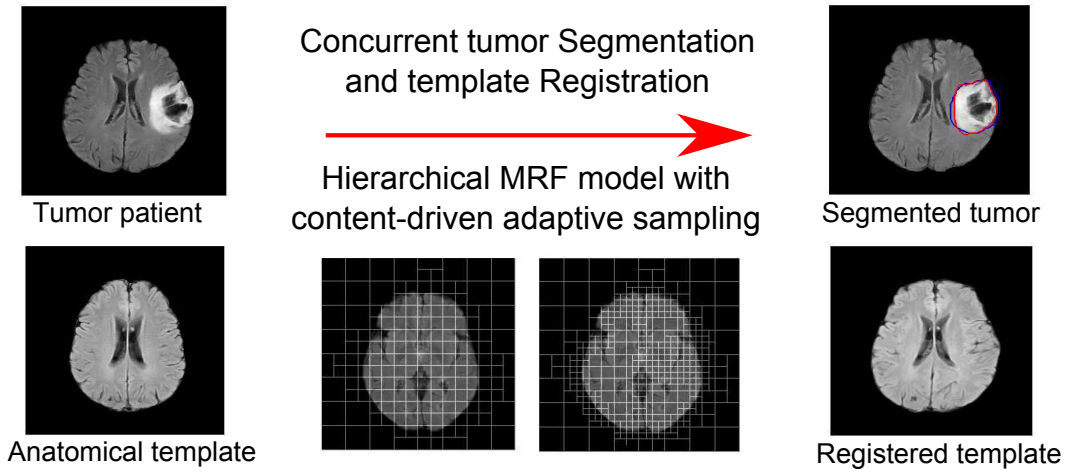


Figure 4.1: *Overview of the Joint Registration and Segmentation method*

is thus treated sequentially. Such methods do not exploit the inherent link between segmentation and registration, where the segmentation benefits from a good registration (atlas anatomical information) and the registration benefits from the segmentation by taking into account the area without correspondences. This chapter presents a concurrent registration and segmentation framework where the segmentation map and deformation field are estimated simultaneously leading to better solutions for both problems. A Markov Random Field model is adopted, seeking to assign to each graph node a label that corresponds to a displacement and a segmentation label. Statistical classification is used to provide tumor vs background prior likelihoods along with conventional registration similarity metrics. In the missing correspondences area (tumor), the registration term is relaxed and the area is implicitly deformed by interpolation.

The remainder of this chapter is organized as follows: section 4.2 present the related work on brain registration in the presence of a pathology and concurrent registration and segmentation. Section 4.3 introduces the joint MRF framework while section 4.4 present a stochastic approach to account for the main drawback of discrete MRF based registration, which is the trade off between complexity and sampling precision. Experimental validation will be part of section 4.4 and a discussion concludes this chapter. A short graphical overview of the method is shown in Fig. [4.1]

4.2 Related work

The following notations will be used in this section: I is the fixed or target image and J is the moving or source image defined on a domain Ω. $\mathcal{T}$ is the transformation that maps the source image to the target image.

4.2.1 Registration with Missing Correspondences

For the task of non rigid registration in the presence of brain tumors, we can divide prior art into two groups of methods. The first class of methods masks the pathology and does not take it into account during the estimation of the matching criterion. The deformation in and around the tumor is carried out by interpolation via a regularization term. The second group simulate a tumor in the healthy atlas to create correspondences. This is typically designed as an *atlas seeding* process where the origin of the tumor in the atlas (the seed) is determined, followed by the growth of the artificial tumor carried out using growth models.

4.2.1.1 Pathology Masking

A simple approach for pathology masking was proposed in [Brett 2001] and tested on a variety of brain lesions (infarction, metastasis and aneurysm). They used the registration method proposed in [Friston 1995] (distributed in the Statistical Parametric Mapping (SPM) package). An affine registration is followed by a parametric non linear registration. The transformation model is based on a discrete cosine transform (the basis functions are cosine function) and while the Sum of Square Differences was used as matching criterion. The lesion is manually segmented and its corresponding area is masked so that it is not taken into account for the evaluation of the optimal displacement.

In the work of Stefanescu et al. [Stefanescu 2004], the registration problem was designed via two regularization processes. The following objective function is to be minimized:

$$E = SSD(I, J \circ \mathcal{T}) + \beta \int_\Omega k(\mathbf{x}) \left\| \nabla \frac{\partial \mathcal{T}}{\partial t} \right\|^2 + \gamma \int_\Omega D(\mathbf{x}) \|\nabla \mathcal{T}(\mathbf{x})\|^2 \, dx \tag{4.1}$$

The first term is the similarity criterion, defined as the Sum of Square Differences (SSD). The second term is a fluid regularization of the deformation field that allows large local displacements, while the third term is an elastic regularization which role is to preserve anatomical coherence. The space varying weighting parameter $D(\mathbf{x})$ varies with respect to the anatomical tissues deformability, simulating the behavior of a biomechanical model. The parameter that accounts for the presence of the tumor is the fluid regularization weighting parameter $k(\mathbf{x}) \in [0, 1]$. This parameter is a confidence score on the reliability of the matching criterion. If $k(\mathbf{x})$ is close to zero, the smoothness of the deformation is heavily imposed and almost interpolated from neighboring values. In healthy structures, the value of k depends on the image gradient. It is set to zero in the tumor area, and therefore the deformation field is interpolated. The optimal displacement is recovered through a four steps iterative process. First the velocity field $v(\mathbf{x})$ is recovered by gradient descent of the SSD criterion (which amounts to approximating the demons algorithm forces [Pennec 1999]). Second, a weighted fluid regularization of this field is performed as $\frac{\partial v}{\partial t}(\mathbf{x}) = k(\mathbf{x})\Delta v(\mathbf{x})$. Eventually, the deformation field is updated $\mathcal{T}^{t+1} = \mathcal{T}^t \circ v$ and smoothed by elastic regularization $\frac{\partial \mathcal{T}}{\partial t}(\mathbf{x}) = \nabla(D(\mathbf{x})\nabla\mathcal{T}(\mathbf{x}))$. An extension of the method was presented in [Commowick 2005] to construct an anisotropic weighting field $D(\mathbf{x})$ based on image statistics.

The method was more recently adapted to the diffeomorphic demons framework [Vercauteren 2008] in [Lamecker 2010]. Thirion's algorithm was cast as a well posed energy minimization problem via the introduction of hidden variables called *correspondences*. The idea is that instead of requiring point to point correspondences to be exact, local errors are allowed [Cachier 2003]. This translates in the following energy:

$$E(C,\mathcal{T}) = \frac{1}{\sigma_i^2} Sim(I, J \circ C) + \frac{1}{\sigma_c^2} dist(\mathcal{T}, C) + \frac{1}{\sigma_r^2} Reg(\mathcal{T}) \tag{4.2}$$

Where σ_c and σ_r are weights depending on the spatial uncertainty of the correspondences and on the irregularity of the deformation respectively. Typically, the different terms are defined as $Sim(I, J \circ C) = \|I - J \circ C\|^2$, $dist(\mathcal{T}, C) = \|C - \mathcal{T}\|^2$ and $Reg(\mathcal{T}) = \|\nabla \mathcal{T}\|^2$. This formulation enables to decouple the problem into two iterative optimization steps:

$$\begin{aligned} &\text{Correspondence step:} && C^t = \underset{C}{\operatorname{argmin}} \ \frac{1}{\sigma_i^2} Sim(I, J \circ C) + \frac{1}{\sigma_c^2} dist(\mathcal{T}^t, C) \\ &\text{Regularization step:} && \mathcal{T}^{t+1} = \underset{\mathcal{T}}{\operatorname{argmin}} \ \frac{1}{\sigma_c^2} dist(\mathcal{T}, C^t) + \frac{1}{\sigma_r^2} Reg(\mathcal{T}) \end{aligned} \tag{4.3}$$

In order to deal with the presence of the pathology, [Lamecker 2010] introduced a third step called the *inpainting step*. This step overwrites the computed displacement field in the tumor area with the solution of the Laplace equation $\Delta \mathcal{T} = 0$. Since no information is given about the image in this area, the deformation is simply interpolated from the boundary of the tumor.

4.2.1.2 Atlas Seeding

Aside from pathology masking, a series of methods specifically designed for brain tumors have been developed. The common idea is to simulate the presence and growth of the tumor in the healthy atlas. A simple method was proposed by Dawant et al. [Dawant 1999]. First, a strongly regularized demons registration brings the atlas and patient images into approximate alignment. The strong regularization limits the amount of displacements allowed which preserves the brain anatomy and prevents errors due to the presence of the tumor. Tumor voxels in the patient's image are manually segmented and assigned a constant intensity value. A subset of the synthetic lesion voxels (selected by erosion of the segmentation map) are then artificially introduced in the healthy atlas (the atlas is said to be *seeded*). A second weakly regularized registration is carried out where the seed will trigger demon forces that will simulate the tumor growth in the atlas to match the patient's tumor. This implies however a sizable seed that can mask key anatomical information in the atlas.

The work of [Bach Cuadra 2004, Pollo 2005] was a natural extension of this seeding approach. Automatic segmentation of the tumor is carried out using the statistical classification based ATM-SVC algorithm [Warfield 2000, Kaus 2001] (see section 2.5.3 for an overview of the method). Here, the seed (a single voxel in the atlas) is manually chosen by an expert. The registration method follows the demons algorithm but treats the voxels differently if they are inside or outside the tumor.

The displacement field inside the tumor follow a growth model that assumes a radial growth from the seed. The classical demon forces are computed outside the tumor. The displacement field is adaptively regularized using a Gaussian filter of variable standard deviation σ. The tumor area is not regularized ($\sigma = 0$), a small value is used at the tumor boundary ($\sigma = 0.5$) and smaller deformations are allowed in the rest of the brain ($\sigma = 0.8$). A variant was proposed later in [Bach Cuadra 2006] using a Mutual Information flow based registration scheme in the healthy region. The drawbacks of the method include expert seeding and a limitation to homogeneous and uniformly growing meningiomas in terms of growth model and automatic segmentation.

A second group of methods relied on biomechanical models to simulate the deformations induced by the tumor. The work of Kyriacou et al. [Kyriacou 1999] was the first using biomechanical modeling for atlas to patient registration in the presence of a tumor. A Finite Element (FE) model is used to model the brain's deformations. The FE method is a numerical approach to approximate the solution of partial differential equations with boundary conditions by discrete meshing of the domain. Here, the brain boundaries are defined manually and used to construct a triangular mesh of the brain. The different brain structures are modeled as elastic materials with tissue specific parameters. Similarly to [Bach Cuadra 2004], the tumor growth is assumed to be radial and uniform. Given a circular seed of diameter D_s, the grown tumor would have a diameter of $D_t = D_s(e_0 + 1)$, where e_0 is a parameter that specifies the growth rate. Because of boundary constraints defined from the surrounding tissues, the resulting growth of the tumor will tend not to be uniform. First the model shrinks the tumor to an almost inexistent tumor. A normal registration is then carried out between the healthy atlas and simulated healthy patient. The third step simulates the growth of the tumor in the registered atlas via a regression procedure to estimate the seed location in the atlas. This methods inherits a high computational cost and was limited to 2D models.

Mohamed et al. [Mohamed 2006] proposed a registration framework that decomposed the deformation field between the healthy atlas and diseased patient into a tumor induced deformation (the mass-effect) and an inter-subject anatomical differences deformation. They used a more sophisticated finite elements biomechanical model of tumor growth [Mohamed 2005]. Tumor and edema are modeled as concentric spheres following the idea of [Kyriacou 1999]. The mass effect induced by the tumor is modeled by a constant outwards pressure P from the tumor's boundary, while an isotropic expansion in white matter is assumed for the edema. In order to reduce the computational complexity inherited from the biomechanical model, a PCA based statistical model was used to learn the deformations induced by the tumors for different subjects and parameters (tumor and edema center and radii, and pressure P). Given such a model, an initial guess of the deformation field obtained by normal registration is projected onto the tumor induced deformation space. This enables the estimation of the optimal tumor growth parameters via a maximum likelihood framework and the simulation of the tumor in the atlas. The final step registers the tumor bearing atlas to the patient.

Zacharaki et al. [Zacharaki 2008] built upon this work using a more elaborated registration scheme and method for the evaluation of the tumor growth parameter. Their framework, called ORBIT, adapts the HAMMER algorithm [Shen 2002] to the presence of tumors. The HAMMER algorithm is a hierarchical registration framework that was proposed in order to reduce the ambi-

guities of intensity matching driving the registration process to local minima. The main idea was to compute a similarity metric between *attribute vectors* instead of relying solely on the intensity values. The attribute vectors are computed at each image voxel and consist of edge information, normalized intensities and tissue specific geometric moments invariants (computed on a spherical neighborhood at different scales). In ORBIT, two tumor based attributes were introduced: the signed distance from the tumor boundary and the angular location with respect to its center. The similarity measure between two voxels is defined as a weighted sum of a criterion matching the healthy structures Sim_B and a tumor criterion Sim_T that matches the tumors' geometry:

$$Sim(\mathbf{x},\mathbf{y}) = (1 - w(\mathbf{x},\mathbf{y}))Sim_B(\mathbf{x},\mathbf{y}) + w(\mathbf{x},\mathbf{y})Sim_T(\mathbf{x},\mathbf{y}) \tag{4.4}$$

$w(\mathbf{x},\mathbf{y})$ is a weighting factor that is equal to one in the tumor area and decreases proportionally with respect to the distances to both tumors (one simulated in the atlas and the target's image tumor). Sim_B matches the healthy attributes (edges, intensity, geometric moments) while the tumor attributes are matched by Sim_T. The registration process starts by selecting the voxels with the most salient features (the driving voxels) that are then hierarchically enriched by new ones to increase the registration accuracy. The biomechanical model parameters are estimated by minimizing an objective function based on the assumption that a poor modeling of the tumor will produce important and unrealistic deformations. The objective function evaluates the quality of the obtained registration using 3 terms: (i) the segmentation overlap of the registered images, (ii) the similarity metric, and (iii) the laplacian of the deformation field which measures its smoothness. A new registration process is then carried out with the optimally estimated parameters. A summary of the iterative approach can be seen in Fig.[4.2]. The method was improved in [Zacharaki 2009] with a computationally efficient biomechanical model [Hogea 2007] that does not require statistical training.

The appeal of growth models is undeniable, as they offer means to obtain anatomically realistic deformations induced by the tumor. However, the quality of the registration depends on the accuracy of the model and of its parameters. Simple radial growths are far from adapted to the many types of tumors while the biomechanical models suffer from their computational cost in modeling and seeking the optimal parameters. They are also prone to errors when dealing with infiltrative tumors, mistaking infiltrated tissue with new tissue. Despite its simplicity, pathology masking is more adapted to the diffusively infiltrative tumors that are DLGGs and could also be used in other clinical contexts since no assumption is made on the pathology nor its progression. A common drawback for both approaches is their dependency on an accurate pre-segmentation of the tumor.

4.2.2 Joint Registration and Segmentation

4.2.2.1 Healthy Volumes

Coupling registration and segmentation is not a new problem and several attempts have been made in the context of healthy structures. The key idea of these methods is that the structures to be segmented are present in both images. Yezzi et al. [Yezzi 2003] proposed an active contour based

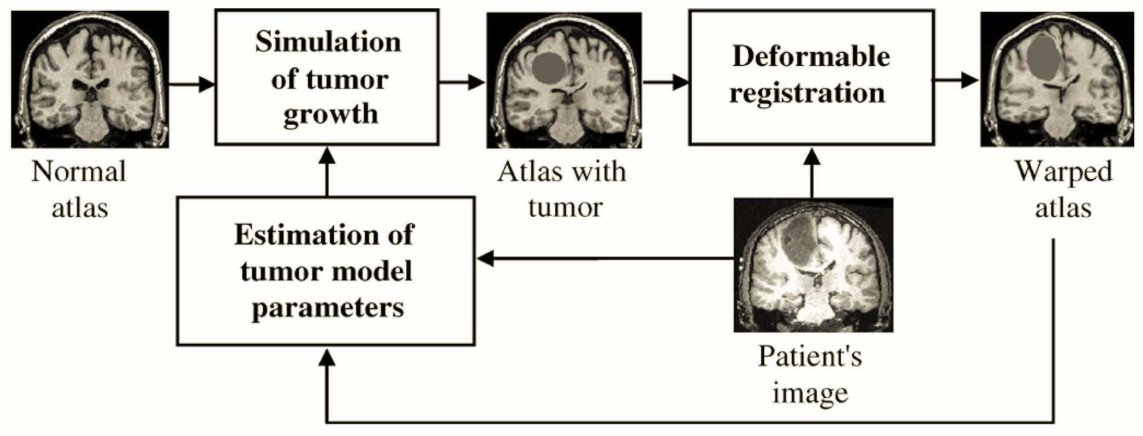


Figure 4.2: *Summary of different steps of the ORBIT algorithm. Figure extracted from [Zacharaki 2008].*

framework where the goal is to segment the same object in two different images I and J. The goal is to find the contour $\mathcal{C}_I$ of the object in I and the transformation $\mathcal{T}$ that maps I to J so that $\mathcal{C}_J = \mathcal{T}(\mathcal{C}_I)$. An active contour energy is defined on each image as a sum of region terms.

$$E_I = \int_{InCurve} f_{in}(\mathcal{C}_I) + \int_{OutsideCurve} f_{out}(\mathcal{C}_I) \tag{4.5}$$

where the functional f_{in} and f_{out} are the difference between the observed intensity and mean intensity in their corresponding region. A rather similar concept was used in [Paragios 2005] where the aim was joint segmentation and sense optical flow estimation. The link between contours $\mathcal{C}_I$ and $\mathcal{C}_J$ enables to define a coupled objective function as the sum of both energies $E(\mathcal{T}, \mathcal{C}) = E_I(\mathcal{C}_I) + E_J(\mathcal{T}(\mathcal{C}_I))$. The type of transformations where limited to rigid and affine transformations that have a finite number of parameters to evaluate. This idea was extended in the context of dense optical flow in [Karantzalos 2005].

Wyatt and Noble [Wyatt 2003] cast the problem of joint segmentation and rigid registration as a MAP estimation one regarding the optimal segmentation maps $\mathcal{S}_I, \mathcal{S}_J$ and deformation field $\mathcal{T}$. Assuming independence between observations I and J, the joint probability is written as:

$$P(\mathcal{S}_I, \mathcal{S}_J, \mathcal{T}|I, J) \propto P(I|\mathcal{S}_I)P(\mathcal{S}_I)P(J|\mathcal{S}_J)P(\mathcal{S}_J)P(\mathcal{T}) \tag{4.6}$$

A MAP-MRF formulation is used for the segmentation terms. The pairwise term follows the Potts model while the likelihood terms are mixtures of Gaussians. The rigid registration uses the mutual information similarity determined according to the joint class histogram obtained from the segmentation. The method alternates between estimating the optimal segmentation fields for both images by maximizing the segmentation posterior probability and estimating the registration parameters by minimizing the similarity measure.

Ashburner et al. [Ashburner 2005] presented a probabilistic framework for simultaneous tissue classification, atlas registration and bias field correction. The model is based on classifying the image using a mixture of Gaussians. Bias field correction is included in the model by locally adjusting the mean and standard deviation values of the Gaussian distributions. The local variations are governed by an additional parameter β that accounts for smooth intensity variations across the image. Spatial priors are introduced as the ICBM probabilistic atlas values. After an initial affine registration, the matching between the atlas and the image to be segmented is refined by the introduction of a parameter α that allows local deformation of the atlas. Instead of weighting the Gaussians by a constant parameter $P(S(\mathbf{x}) = k|\gamma) = \gamma_k$, they are weighted by a spatially varying prior:

$$P(S(\mathbf{x}) = k|\gamma, \alpha) = \frac{\gamma_k \Pi_k(\alpha(\mathbf{x}))}{\sum_{i=1}^{K} \gamma_i \Pi_i(\alpha(\mathbf{x}))} \tag{4.7}$$

where Π_k is the atlas probability value for class k and K the number of classes. The transformation is parametrized as a linear combination of cosine transforms. Smoothing priors for the bias field vector and deformation field are added as multivariate Gaussians of zero mean. The optimal parameters are obtained by minimizing the following objective function using the ICM algorithm [Besag 1986]:

$$E(\Phi, \beta, \alpha) = -\sum_{i=1}^{K} \log\left(\mathcal{N}\left(\frac{\mu_k}{\rho(\beta)}, \frac{\sigma_k}{\rho(\beta)}\right) P(S(\mathbf{x}) = k|\gamma, \alpha)\right) - \log P(\alpha) - \log P(\beta) \tag{4.8}$$

where Φ corresponds to the set of Gaussian parameters.

An Expectation Maximization framework was proposed by Pohl et al. [Pohl 2006] for the estimation of the transformation, structures segmentation and bias field correction parameters. The method follows the segmentation and bias field estimation method of Wells et al. [Wells III 1996b] but integrates affine registration and atlas prior information. The E-step estimates the classes posterior probabilities based on the current bias field and aligned spatial priors (probabilistic atlas maps). The M-step updates the registration and bias field parameters based on the classes probabilities. The registration parameters are computed based on the current segmentation and minimize the Kullback Leibler divergence between the computed segmentation posterior probabilities and the probabilistic atlas prior probabilities. In other words, we seek the best possible agreement between both probability distributions.

Last but not least, Mahapatra and Sun [Mahapatra 2012] introduced a 2D MRF based concurrent registration and segmentation method. Each voxel is assigned a label that corresponds to a displacement vector and segmentation label. The unary term is the sum of a conventional registration similarity measure and a term that computes the agreement between the source and target images intensity based classes likelihoods. The pairwise term ensures smoothness of the deformation field by forcing pixels of the same class to have similar displacements and limiting the maximum displacement between voxels of different classes. The MRF is optimized using graph cuts.

4.2.2.2 Tumor Bearing Images

Joint segmentation and registration methods for healthy volumes exploit the fact that the same objects are present in both volumes. The problem becomes far more difficult in the presence of a tumor where the object sought to segment is absent in one of the volumes. Gooya et al. [Gooya 2012] proposed a Bayesian approach to tackle it, inspired from the work of [Zacharaki 2009] and [Pohl 2006]. They simulate a tumor in the atlas using a diffusion-reaction growth model coupled with an elastic modeling of the brain to simulate the mass effect [Hogea 2008]. An example of modified atlas is shown in Fig. 4.3. The optimal registration and growth model parameters as well as segmentation classes are inferred using the EM algorithm guided by the modified atlas. This joint segmentation and registration work is an extension of their growth model-EM registration framework [Gooya 2011], where only the registration and growth model parameters were inferred through SVM classification. Here, they work on multispectral images (T1, T1CE, T2, and FLAIR) and model the different classes as multivariate Gaussian distributions weighted by the atlas probability maps. The EM E-step computes the classification posterior probability given the current estimates of the Gaussian, growth model and registration parameters that are updated in the M-step.The method suffers from the drawbacks of growth model approaches, that is computational cost and lack of adaptability.

Related work in the context of surgical resection will be introduced in appendix A.

4.3 Concurrent Deformable Registration and Tumor Segmentation

Efficient coupled registration and segmentation frameworks, should inherit reasonable computational complexity while being modular with respect to the different components of the algorithm (similarity criterion for registration, tumor detection method, and even clinical context). In such a context, the model of choice was an MRF model that coupled with discrete optimization algorithms inherits such properties. In order to account for the presence of the tumor, we selected pathology masking instead of growth models since it didn't require prior knowledge on the evolution of the tumor and could easily be introduced in an MRF model. Furthermore, the deformation induced by DLGGs tends to be limited due to their infiltrative nature. Contrarily to pathology masking methods that require an initial segmentation and can introduce a bias in the registration process, in our approach the mask is progressively determined during registration and becomes more and more precise as the images get registered.

4.3.1 Discrete MRF Model for Deformable Registration

The formulation of the registration problem as an MRF optimization problem was introduced in [Glocker 2008a]. The problem's complexity was remarkably reduced by the use of a parametric deformation model as well as a discretized solution space. Consider a moving image A defined on a domain Ω and a fixed image I. We seek the transformation $\mathcal{T}$ that will map the moving

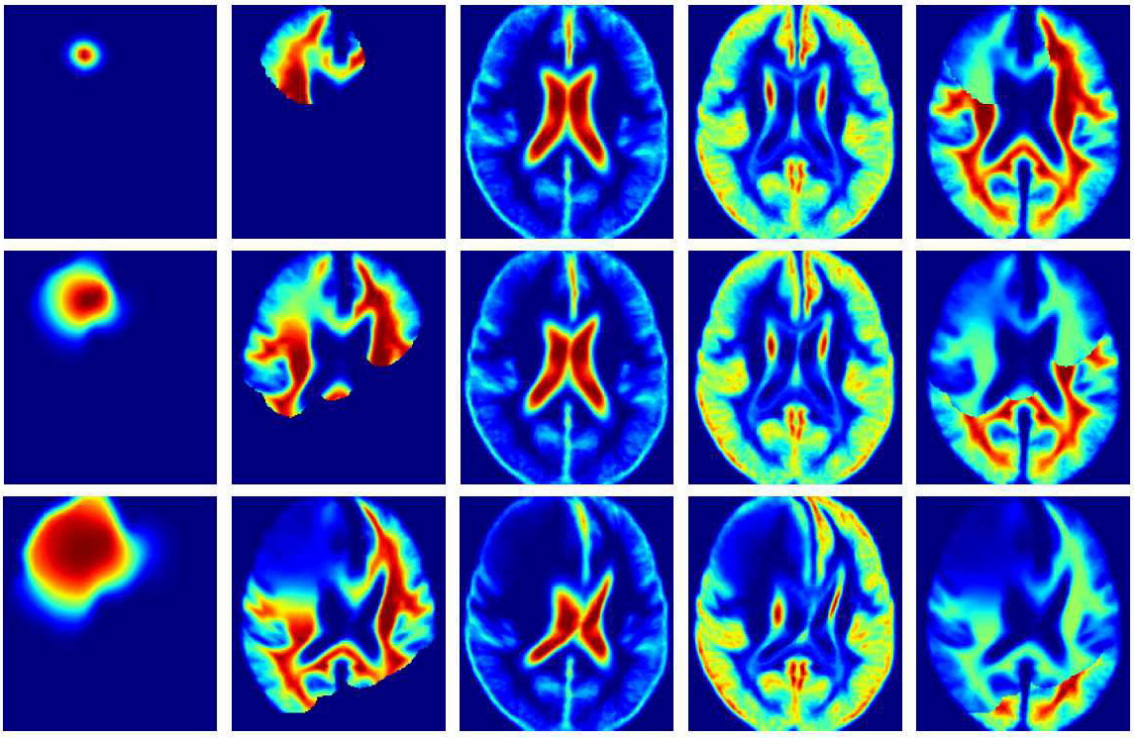


Figure 4.3: *Modified ICBM probabilistic atlas by the tumor growth model. From left to right: tumor, edema, CSF, gray matter and white matter probabilities. Figure extracted from [Gooya 2012]*

image to the fixed image. In the context of parametric transformation models, Free Form Deformation models [Sederberg 1986] assume a uniform sparse grid superimposed to the moving volume [Rueckert 1999]. The grid's control points constitute a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the nodes correspond to the control points and the edges constitute a neighborhood system $\mathcal{N}$. We define a predefined discrete displacement set $\Delta = \{\mathbf{d^1}, ..., \mathbf{d^n}\}$ and a set of labels $\mathcal{L} = \{1, ..., n\}$. Each graph node $\mathbf{p}$ is to be assigned a label $l_p \in \mathcal{L}$ that is associated to a displacement $\mathbf{d^{l_p}} \in \Delta$. The complete deformation field is then estimated by interpolation:

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \sum_{\mathbf{p} \in \mathcal{G}} \eta(\|\mathbf{x} - \mathbf{p}\|)\mathbf{d^{l_p}} \tag{4.9}$$

where $\eta(\|\mathbf{x} - \mathbf{p}\|)$ is the projection function that describes the influence of control point $\mathbf{p}$ over image voxel $\mathbf{x}$. When the FFD model is adopted, the projection function is defined as cubic B-splines.

Let us recall that the form of a deformable registration energy to be minimized is the following:

$$E(\mathcal{T}) = Sim(I, A \circ \mathcal{T}) + \lambda Reg(\mathcal{T}) \tag{4.10}$$

The optimal labeling $\mathbf{l} = \{l_p, \mathbf{p} \in \mathcal{G}\}$ is recovered by minimizing the MRF energy

$$E(\mathbf{l}) = \sum_{p \in \mathbf{G}} V_p(l_p) + \sum_{p \in \mathcal{G}} \sum_{q \in \mathcal{N}(\mathbf{p})} V_{pq}(l_p, l_q) \tag{4.11}$$

The parallel between an MRF energy and the deformable registration's energy is clearly visible. The pairwise costs play correspond to registration regularization inherent to non-rigid registration methods. The unary costs correspond to the similarity criterion that is to be minimized to match the images. In the work of Glocker et al., it is defined as a point wise criterion:

$$Sim(I, A \circ \mathcal{T}) = \int_{\Omega} Sim(I(\mathbf{x}), A(\mathcal{T}(\mathbf{x}))) \, dx \tag{4.12}$$

Because of the parametric formulation, the energy could be expressed through terms acting on the grid's control points. To this end, we need to back project the voxel wise similarity measure on the control points:

$$Sim(I, A \circ \mathcal{T}) = \frac{1}{|\mathcal{G}|} \sum_{\mathbf{p} \in \mathcal{G}} \int_{\Omega} \hat{\eta}(\|\mathbf{x} - \mathbf{p}\|) Sim(I(\mathbf{x}), A(\mathcal{T}(\mathbf{x})) \, dx \tag{4.13}$$

where $\hat{\eta}(\|\mathbf{x} - \mathbf{p}\|)$ is the back projection function that describes the influence of voxel $\mathbf{x}$ on control point $\mathbf{p}$. For a voxel wise similarity measure, it is defined as:

$$\hat{\eta}(\|\mathbf{x} - \mathbf{p}\|) = \frac{\eta(\|\mathbf{x} - \mathbf{p}\|)}{\int_{\Omega} \eta(\|\mathbf{y} - \mathbf{p}\|) dy} \tag{4.14}$$

The unary costs are defined according to the similarity criterion as:

$$V_p(l_p) \approx \frac{1}{|\mathcal{G}|} \int_{\Omega} \hat{\eta}(\|\mathbf{x} - \mathbf{p}\|) Sim(I(\mathbf{x}), A(\mathcal{T}(\mathbf{x}))) \, dx \tag{4.15}$$

A key hypothesis of the MRF formulation is that the unary term $V_p(l_p)$ depends only on its corresponding node $\mathbf{p}$. However, the formulation of the unary cost proposed in equation 4.15 does depend on the neighboring nodes because of the interpolation term in $\mathcal{T}(\mathbf{x})$. It was shown in [Glocker 2008b] that to preserve independence, the unary costs can approximate the similarity value as:

$$V_p(l_p) = \frac{1}{|\mathcal{G}|} \int_{\Omega} \hat{\eta}(\|\mathbf{x} - \mathbf{p}\|) Sim(I(\mathbf{x}), A(\mathbf{x} + \mathbf{d}^{l_p})) \, dx \tag{4.16}$$

Considering that the fixed volume I features a pathology, the MRF registration method has to be adapted. Following the pathology masking methods, a natural approach would be to ignore the similarity criterion in the tumor area, which will be deformed by interpolation due to the regularization term. Two issues arise from such an approach. First, it is required that the tumor's position is known which implies the need for an initial segmentation. Second, masking the pathology can introduce an initial bias and prevent the correct registration of healthy structures that are close to the tumor. Errors get even more likely when the automatic segmentation of the tumor is not that precise.

4.3.2 Coupled Tumor Segmentation and Registration

Consider a moving image or reference brain atlas A defined on a domain Ω and a fixed tumor bearing brain image I. The objective here is twofold: register the healthy atlas to the diseased subject and segment the tumor. Given the strong dependencies between the two problems, we aim at simultaneously finding the deformation field $\mathcal{T}$ and the segmentation map $\mathcal{S}$. Aside from increasing the registration quality, coupling segmentation with registration also benefits the segmentation output as it introduces anatomical knowledge in the segmentation process and reduces the risks of false positives.

We consider a sparse grid $\mathcal{G}$ superimposed to the atlas, a predefined set of n discrete displacements $\Delta = \{\mathbf{d^1}, .., \mathbf{d^n}\}$ and a discrete label set $\mathcal{L}_c = \{1, .., 2n\}$. The problem of concurrent tumor segmentation and registration is converted into a discrete labeling problem. Assigning a label l_p to a grid node $\mathbf{p}$ corresponds to assigning a pair segmentation/displacement $\{\mathbf{d^{l_p}}, s^{l_p}\} \in \Delta \times \{0, 1\}$ where :

$$\begin{aligned} \mathbf{d^{l_p}} &= \mathbf{d^k} \text{ where } k = l_p \pmod n \\ s^{l_p} &= \left\lceil \frac{l_p}{n} \right\rceil \end{aligned} \tag{4.17}$$

As a result, when a grid node is assigned a label, it is simultaneously assigned a displacement (selected from the predefined discrete set) and labeled as a tumor or background node. The displacements and classification decisions are then propagated on the whole volume by interpolation:

$$\begin{aligned} \mathcal{T}(\mathbf{x}) &= \mathbf{x} + \sum_{\mathbf{p}\in\mathcal{G}} \eta(\|\mathbf{x} - \mathbf{p}\|)\mathbf{d^{l_p}} \\ \mathcal{S}(\mathbf{x}) &= \sum_{\mathbf{p}\in\mathcal{G}} \eta(\|\mathbf{x} - \mathbf{p}\|)s^{l_p} \end{aligned} \tag{4.18}$$

The MRF energy to minimize is designed as an association of segmentation and registration terms:

$$E_c = \sum_{\mathbf{p}\in\mathcal{G}} \underbrace{\alpha V_R(l_p) + (1-\alpha)V_S(l_p)}_{\substack{\text{Registration and segmentation} \\ \text{unary terms}}} + \sum_{\mathbf{p}\in\mathcal{G}} \sum_{\mathbf{q}\in\mathcal{N}(\mathbf{p})} \alpha \underbrace{V_{pq,R}(l_p, l_q)}_{\substack{\text{Registration} \\ \text{pairwise cost}}} + (1-\alpha) \underbrace{V_{pq,S}(l_p, l_q)}_{\substack{\text{Segmentation} \\ \text{pairwise cost}}} \tag{4.19}$$

Where α is a parameter that balances the importance of the registration and segmentation term. The focus is on registration when α is high, and on segmentation when α is low. Linear programming is used to recovered the optimal labeling via the Fast-PD algorithm [Komodakis 2008b].

4.3.2.1 Unary terms

The registration term seeks for optimal correspondences between the fixed and moving image. Correspondences are sought in the healthy areas by computing a similarity measure $Sim(.)$ between the fixed and deformed image. However, they do not exist in the tumor area (characterized by

$s^{l_p} = 1$) where the similarity measure becomes unreliable. We use a constant cost C_{tm} instead whose role is two-fold: first it provides a unary cost independent of the displacement so that the deformation is solely guided by the pairwise regularization term. Second, It identifies regions that are potentially tumors due to a high dissimilarity between the two images:

$$V_R(l_p) = \int_\Omega \hat{\eta}(\|\mathbf{x}-\mathbf{p}\|) \left(\underbrace{s^{l_p} C_{tm}}_{\substack{\text{Tumor case:} \\ \text{Constant cost}}} + \underbrace{(1-s^{l_p}) Sim\left(I(\mathbf{x}), A(\mathbf{x}+d^{l_p})\right)}_{\text{Background case: Similarity criterion}} \right) dx \tag{4.20}$$

The segmentation term corresponds to the prior likelihood probability of the fixed volume I: $p(I(\mathbf{x})|s^{l_p})$. It refers to voxel wise classes probabilities computed using standard classification techniques. This term ensures that the nodes with a high probability of being tumor are labeled accordingly. Since the node is segmented and displaced simultaneously, the segmentation term is dependent on the node's displacement $\mathbf{d^{l_p}}$:

$$V_S(l_p) = \int_\Omega \hat{\eta}(\|\mathbf{x}-\mathbf{p}\|) \left(-s^{l_p} \log p\left(I(\mathbf{x}+\mathbf{d^{l_p}})|s^{l_p}=1\right) - (1-s^{l_p}) \log p\left(I(\mathbf{x}+\mathbf{d^{l_p}})|s^{l_p}=0\right)\right) dx \tag{4.21}$$

For both unary terms, the back projection function $\hat{\eta}(.)$ is used to back project the voxel wise information (class probabilities and similarity criterion) on the control points.

4.3.2.2 Pairwise terms

The role of the pairwise terms is to enforce smoothness of the segmentation and deformation field. If the resolution of the grid is low (large spacing between the nodes), it makes little sense to impose similar labels to the neighboring nodes. Furthermore, cubic B-spline interpolation on a large resolution grid produces a satisfactory smoothing effect. The pairwise costs are therefore strengthened as the resolution decreases. The segmentation pairwise cost is designed as a simple Potts model weighted by the distance between the nodes:

$$V_{pq,S}(l_p, l_q) = \frac{|s^{l_p} - s^{l_q}|}{\|\mathbf{p}-\mathbf{q}\|} \tag{4.22}$$

The registration pairwise cost is defined in a similar fashion and imposes local consistency of the displacement field:

$$V_{pq,reg}(l_p, l_q) = \lambda \frac{\left(\mathbf{d^{l_p}} - \mathbf{d^{l_q}}\right)^2}{\|\mathbf{p}-\mathbf{q}\|} \tag{4.23}$$

where λ describes the influence of the registration regularization. The registration pairwise cost plays a key role in the tumor area as it drives the deformation (due to the constant registration unary cost) and determines the amount of admissible deformation. In healthy areas, λ should be high enough to preserve the anatomical structures. It should however be weakened around the tumor area where strong deformations can occur due to the tumor's growth.

4.3.2.3 Incremental implementation

The main limitation of such discrete models resides in the definition of the displacement search space and of the grid's resolution. On one hand, we want the algorithm to handle large and yet precise displacements which requires a large and thorough displacement set as well as a important grid resolution. On the other hand, they are both to be constrained in order to maintain the computational burden manageable. Adding segmentation labels in the process increases the difficulty as almost voxel wise labelings are required, especially given the arbitrary shapes of the tumor.

Those drawbacks can be handled by adopting a coarse to fine hierarchical approach. This enables to recover large deformation (low grid resolution) as well as precise displacements and segmentation at high resolution while defining small discrete displacement sets [Glocker 2008a]. Additionally, the propagation of segmentation decisions from low resolution to high resolution makes the segmentation less sensitive to noise and the tumor's heterogeneous appearance. In this setting, grids of increasing resolutions are considered: $\{\mathcal{G}_r, r \in [1, L]\}$, where L corresponds to the number of resolution levels. For increased precision, several iterations (MRF optimizations) are performed at each grid level r during which the displacement set Δ_t is progressively refined around the optimal displacement obtained at iteration $t-1$. At each iteration, the updated MRF energy is computed between the fixed image and the deformed image $A \circ \mathcal{T}^{t-1}$ on the basis of the deformation field estimated at the previous iteration:

$$\mathbf{l}^t_{opt} = \underset{\mathbf{l}}{\operatorname{argmin}}\, E^t_c(\mathbf{l}\,|I, A \circ \mathcal{T}^{t-1}) \tag{4.24}$$

The newly obtained deformation field after optimization is then composed with $\mathcal{T}^{t-1}$.

Three challenges arise from such a formulation for concurrent registration and segmentation. First, the segmentation decisions obtained at resolution $r-1$ should be propagated to the next level. The idea is that images at a low resolution should not suffer from local non homogeneity in tumor appearances resulting in a smoother segmentation than what would be obtained for a voxel wise segmentation. The finer the resolution becomes, the more focus should be paid on the tumor's boundaries. In [Parisot 2012b], we introduced into the MRF energy a cost penalizing classification decisions that are not in accordance with the segmentation map of the previous level:

$$V^t_{pen}(l_p) = \int_\Omega \bar{\eta}(\|\mathbf{x}-\mathbf{p}\|) s^{l_p} \mathcal{S}^{t-1}(\mathbf{x}) \frac{2}{\exp(r)}\, dx \tag{4.25}$$

However, such a cost is manually set and has to be adapted to the clinical context and potentially, to each patient. Furthermore, erroneous segmentation can be propagated from one level to the next that is not a desired behavior since only the most certain labels should be propagated.

Second, the grid resolution levels have to be specified: the first and last resolution as well as the refinement process from one level to the next. This is the key aspect of the incremental approach in terms of efficiency and quality of registration/segmentation. While fine resolutions could lead to precise detection of the tumor's boundary and displacements, they also suffer from a decrease of robustness and increase substantially the computational burden. Typically, the grid is refined

by doubling the number of nodes at each level, which limits the final level's resolution. Ideally, high resolutions should only be considered in areas with discontinuities. This is the idea behind the work of Shi et al. [Shi 2012], who proposed the Sparse Free Form Deformations, based on the assumption that it is possible to embed a realistic transformation in a sparse representation using a multi level FFD framework. Indeed, the B-spline coefficients (control points displacements) are often close to zero for the majority of control points. Given a set of m FFD grids of different resolutions, they aim at simultaneously estimating the control points' displacements of the m levels. This is achieved by adding a sparsity constraint on the estimated solution to the registration objective function. This results in a sparse representation of the deformation field across multiple resolutions, where importance is given to control points at different resolutions according to the observed images properties.

The third challenge is the definition and resampling of the displacement set at each iteration. In [Glocker 2008a], the displacement set is defined as an isotropic regular sampling around the control point's initial position. It is refined at each iteration by simply diminishing the maximum displacement allowed. While it allows additional precision, it fails to take into account the local anisotropy of the regions to be registered. An uncertainty based sampling was proposed in [Glocker 2008b] for MRF based registration. The impact of a label swap (different displacement) on the overall energy is evaluated by computing the min-marginal energies [Kohli 2008]. The uncertain directions are estimated as directions where there are small energy variations and the displacements are then sampled according to those directions.

Inspired by this work, we consider local uncertainties to design a content driven adaptive framework that handles the three main issues of the hierarchical approach. The resampling of the displacement sets and grid resolution is guided by the segmentation and registration uncertainties that are computed from the min marginal energies. A segmentation propagation term is also derived from the uncertainties, assuring propagation of reliable segmentations from the lower resolution levels.

4.4 Uncertainty Driven Resampling

4.4.1 Local Registration and Segmentation Uncertainties

The local segmentation and registration uncertainties are estimated using the min-marginal energies. The use of the min-marginal energies of an MRF to compute the likelihood of a label assignment was suggested in [Kohli 2008] for segmentation and adapted later in [Glocker 2008b] for registration. More recently, new methods for estimation of such min-marginals/uncertainties have been proposed [Tarlow 2012]. By definition, a min-marginal is a function that evaluates the minimum value of the MRF energy under different constraints. In order to evaluate the local uncertainties, the influence of a label selection on the overall energy has to be evaluated. At a given iteration t (associated with resolution r), the min marginal marginal energy associated to a control point $\mathbf{p} \in \mathcal{G}_r$ and a label $k \in \mathcal{L}$ computes the minimum value of the energy E_c^t with the value of l_p

fixed to k:

$$\Psi_{\mathbf{p},k,t} = \min_{\mathbf{l}, l_p=k} E_c^t(\mathbf{l}\,|I, A \circ \mathcal{T}^{t-1}) \tag{4.26}$$

Recall that an MRF energy is associated with the probability of a configuration $P(\mathbf{l})$ as:

$$P^t(\mathbf{l}) = \frac{1}{Z}\exp(-E_c^t(\mathbf{l})) \tag{4.27}$$

Where Z is the partition function. The max marginals $\mu_{\mathbf{p},k,t}$ (evaluated via message passing algorithms) give the value of the maximum probability given the constraint $l_p = k$:

$$\mu_{\mathbf{p},k,t} = \max_{\mathbf{l}, l_p=k} P^t(\mathbf{l}) = \frac{1}{Z}\exp(-\Psi_{\mathbf{p},k,t}) \tag{4.28}$$

By computing the min marginal energies for all possible labels, we can evaluate the influence of a label swap on the energy with respect to the optimal minimum energy. A measure of confidence can be derived from the max marginals by normalizing over all possible labels:

$$C_{\mathbf{p},k,t}(\mathbf{l}) = \frac{\mu_{\mathbf{p},k,t}}{\sum_{i\in\mathcal{L}^t}\mu_{\mathbf{p},i,t}} = \frac{\exp(-\Psi_{\mathbf{p},k,t})}{\sum_{i\in\mathcal{L}}\exp(-\Psi_{\mathbf{p},i,t})} \tag{4.29}$$

Note that the computation of the max marginals' partition function Z is unnecessary since it cancels out when the confidence measure is computed.

In the context of concurrent segmentation and registration, the label information is the product of two label spaces, the displacement and the segmentation space. Considering that label k is different from the optimal label l_p^{opt} and that the segmentation label is constant ($s^k = s^{l_p^{opt}}$), the label swap correspond to a local perturbation from the optimal displacement. A substantial energy increase means that the chosen displacement is quite certain in the direction $\frac{\mathbf{d^k}-\mathbf{d_p^{l_p^{opt}}}}{\|\mathbf{d^k}-\mathbf{d_p^{l_p^{opt}}}\|}$ while a weak energy variation implies that the direction is uncertain and that both labels are almost as likely. A registration confidence measure is computed by normalizing the max marginals over all possible displacements, for a given segmentation label:

$$C_R^t(\mathbf{p},k) = \frac{\exp(-\Psi_{\mathbf{p},k,t})}{\sum_{i\in\mathcal{L}, s^i=s^{l_p^{opt}}}\exp(-\Psi_{\mathbf{p},i,t})} \tag{4.30}$$

A segmentation confidence measure can be similarly derived from the max marginals by measuring the variations of the energy when the segmentation label changes $s^k \neq s^{l_p^{opt}}$. The confidence of a segmentation label S can be measured by normalizing over all labels:

$$C_S^t(\mathbf{p},S) = \frac{\sum_{i\in\mathcal{L}, s^i=S}\exp(-\Psi_{\mathbf{p},i,t})}{\sum_{i\in\mathcal{L}}\exp(-\Psi_{\mathbf{p},i,t})} \tag{4.31}$$

4.4.2 Min Marginals Computation using Dynamic Graph Cuts

While using Fast-PD to compute the min-marginals would be the most straightforward method, it would imply performing an optimization for each fixed label which would considerably slow down the algorithm. We compute an approximation of the min marginals using dynamic graph cuts, following the method proposed by [Kohli 2007]. We recall here that graph cuts rewrite the labeling problem as a graph partitioning one. Each node is connected to a source and sink node, each terminal node corresponding to a label. The optimal labeling is found by solving the min-cut/max-flow problem, that is, finding a partition of the graph by removing the edges of minimal weights. This is equivalent to finding the maximum flow from the source edge to the terminal edge with respect to the edges capacities. Dynamic graph cuts minimize similar functions but with slightly different energy terms. The maximum flow computed for the original problem is recorded and used as initialization for the new, slightly modified problem.

The min-marginals $\Psi_{\mathbf{p},k,t}$ are computed from the graph's source/sink (depending on the fixed label) flow potentials that compute the maximum amount of flow that can arrive/leave the node $\mathbf{p}$. [Kohli 2008] show that they are computed as:

$$\Psi_{\mathbf{p},k,t} = f_{max} + f_{\mathbf{p}}^{T(k)} \tag{4.32}$$

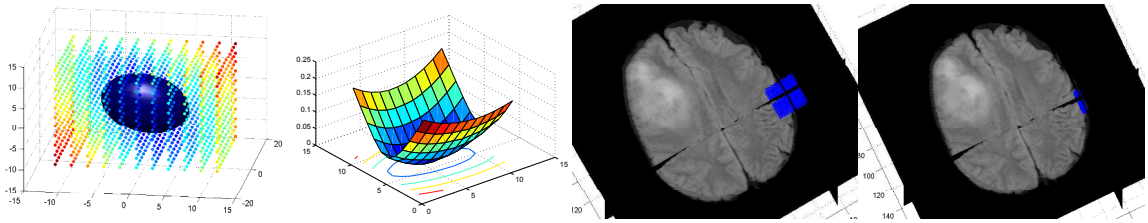
where $T(k)$ is the terminal node corresponding to label k, f_{max} is the maximal flow computed by minimizing the complete MRF energy E_c^t and $f_{\mathbf{p}}^{T(k)}$ is the flow potential. The flow potentials are computed using dynamic graph cuts by adding an edge of infinite cost between node $\mathbf{p}$ and the terminal node $T(k)$ to the residual graph to force the label assignment, and then computing the updated graph's min-cut. While exact solutions can be obtained for binary label sets, the min marginals can only be approximated for mutli-label configurations. They are approximated in the multi-label setting by defining $2n$ binary energies (one per label) where the energy k is computed between the optimal labeling and label k.

4.4.3 Content Driven Adaptive Resampling

Given the segmentation and registration confidence scores, we use this information to design an image driven displacement set and grid resolution dependent on the local image properties. The displacement set construction is guided by the local registration uncertainties, while the grid resolution depends on the homogeneity of the regions covered by the nodes and the associated segmentations confidences.

4.4.3.1 Displacement Sampling

Consider that the current displacement set Δ_t is a dense discretization of the displacement search space defined by a maximum displacement and a sampling rate. Each min-marginal energy $\Psi_{\mathbf{p},k,t}$ (and its corresponding confidence measure) is associated with a spatial location in the Euclidean space centered at node $\mathbf{p}$'s initial position. This spatial location is defined by the position of the



(a) (b) (c) (d)

Figure 4.4: *Registration uncertainty and displacement set resampling for one control point: (a) Min marginal values per displacement label (blue: low, red: high energy) associated covariance matrix centered at the optimal label, (b) Min marginals visualization on a 2D slice, (c) Original isotropic displacement set and (d) Uncertainty driven displacement set, following the brain boundaries.*

node after a displacement of $\mathbf{d}^k$. This yields a confidence map describing which displacements are most likely. We follow the methodology of [Glocker 2008b] and approximate this map as a Gaussian distribution centered at the node's optimal displacement. The distribution's covariance matrix illustrates the directions towards which the displacements are uncertain (small energy variation) and is used to intelligently sample the displacement set Δ_{t+1} according to the covariance's scale and main axes. Note that we do not carry out this process for the tumor label and when the focus is on segmentation (low α) since the registration is mostly driven by the pairwise cost that yields a spherical covariance matrix. An example of adaptive displacement sampling can be seen in Fig.[4.4].

4.4.3.2 Grid Refinement and Segmentation Propagation

We define a uniform grid $\mathcal{G}_{r,max}$ of resolution $M \times N \times P$ and an activation function $\mathcal{A}_r : \mathcal{G}_{r,max} \rightarrow \{0, 1\}$ that describes the resolution of the current grid $\mathcal{G}_r \subset \mathcal{G}_{r,max}$. A node $\mathbf{c_r} \in \mathcal{G}_{r,max}$ is a part of $\mathcal{G}_r$ if $\mathcal{A}_r(\mathbf{c_r}) = 1$. In the remainder of this chapter, the nodes characterized by $\mathcal{A}_r(\mathbf{c_r}) = 1$ will be referred to as *activated nodes*. After optimization, the min-marginal energies are estimated for the activated nodes in $\mathcal{G}_j$.

At the next resolution level $r + 1$, $\mathcal{G}_{r,max}$ is refined as a uniform grid $\mathcal{G}_{r+1,max}$ of resolution $2M - 1 \times 2N - 1 \times 2P - 1$ by splitting the edges of $\mathcal{G}_{r,max}$ in two. The new level's grid resolution $\mathcal{G}_{r+1}$ is determined by activating selected nodes that are likely to increase the precision of the segmentation and registration. There are three possible ways of activating a node at level $r + 1$. First, a node $\mathbf{p} \in \mathcal{G}_{r+1,max}$ will be activated if it has the same spatial coordinates with a node $\mathbf{c_r} \in \mathcal{G}_{r,max}$ that is activated ($\mathcal{A}_r(\mathbf{c_r}) = 1$) at the previous level. Because of the way $\mathcal{G}_{r+1,max}$ is defined with respect to $\mathcal{G}_{r,max}$, all low resolution nodes have a direct spatial correspondent in the higher resolution grid. The activation of the remaining nodes depends on the min-marginal energies evaluated at the low resolution level. A node $\mathbf{p}$ can either be activated for registration

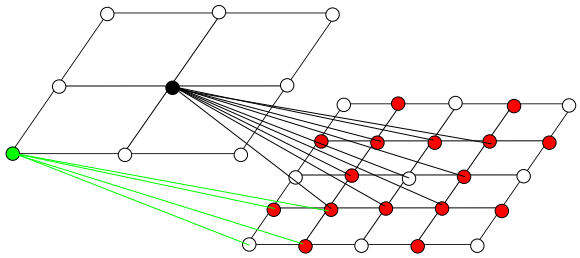


Figure 4.5: *Visual representation of the grid refinement from level j (left) to level j+1 (right). Grid resampling: the nodes that have direct correspondences appear in white, and the new nodes and edges are red. The edges connecting the 2 grids represent the nodes' neighborhood. The grid is shown in 2D for increased visibility.*

or for increased precision of the segmentation. The registration and segmentation activations are governed by two activation terms $\mathcal{A}_R(\mathbf{p})$ and $\mathcal{A}_S(\mathbf{p})$ respectively. In order to propagate the max marginals confidence scores and activation information between levels, we construct an inter-level 3D neighborhood $\mathcal{N}_i(.)$ that connects the node $\mathbf{c_r} \in \mathcal{G}_{r,max}$ to its 27 closest neighbors in $\mathcal{G}_{r+1,max}$, where the proximity is evaluated based on the nodes' spatial coordinates. A 2D representation of the neighborhood can be seen in Fig. [4.5].

The registration activation term ($A_R(\mathbf{p})$) aims at identifying the presence of structures and discontinuities in the region surrounding the node. The idea is that precise displacements (and therefore a higher resolution) are necessary at boundaries while no increased focus is necessary in homogeneous regions. As a result, a node should be activated if it is at the boundary between different structures. If discontinuities are present, strong variations of the min-marginal energies are expected with respect to the displacement label, due to the different structures' properties. The registration activation criterion of a node $\mathbf{p}$ is defined based on the min-marginal energy range of its neighbors in $\mathcal{G}_j$ given their optimal segmentation label:

$$A_R(\mathbf{p}) = H\left(\sum_{\mathbf{c_r}\in\mathcal{N}_i(\mathbf{p})} \frac{1}{N}\left(\max_{k\in\mathcal{L}, s^k=s^{l^{opt}_{cr}}} \Psi_{\mathbf{c_r},k,t} - \min_{k\in\mathcal{L}, s^k=s^{l^{opt}_{cr}}} \Psi_{\mathbf{c_r},k,t}\right) - \mu_r\right) \tag{4.33}$$

where $H(.)$ is the Heaviside step function, N is the number of nodes in the neighborhood of $\mathbf{p}$ and μ_r is the mean energy range over all nodes in $\mathcal{G}_r$. In other words, a node $\mathbf{p}$ will be activated if the mean min-marginal energy range between its neighbors in $\mathcal{G}_r$ is larger than the overall mean energy range μ_r.

Similarly, the segmentation activation term ($A_S(\mathbf{p})$) exploits the information of the min-marginal energies. A confidence score $C^t_S(\mathbf{c_r}, S)$ is computed for each node in $\mathcal{G}_r$ given a label S by normalizing the min marginals. This allows to evaluate how reliable a node's segmentation is. In the case of a binary segmentation, the confidence score can be converted into a node wise uncertainty

measure independent of the label:

$$U^t(\mathbf{c_r}) = 1 - |C_S^t(\mathbf{c_r}, S) - 0.5| \tag{4.34}$$

This term measures the degree of uncertainty associated to the chosen segmentation label. A low value of $U^t(\mathbf{c_r})$ is desirable as it corresponds to highly reliable segmentation decisions. We use this information to propagate the recovered optimal segmentation labels to next levels based on their reliability. This is done by adding an inter-level potential to the MRF energy defined as:

$$V_{ucy}(l_{c_r}^{opt}, l_p) = \begin{cases} 0 & \text{if } s^{l_p} = s^{l_{c_r}^{opt}} \\ -log(U(\mathbf{c_r})) & \text{Otherwise} \end{cases} \tag{4.35}$$

This potential is a penalty cost that penalizes nodes whose the segmentation labels do not agree with their neighbors'. The strength of the penalty depends on the confidence score of the estimated segmentations.

The segmentation activation criterion can easily be derived from this propagation potential: nodes should be activated if their inter-level penalty is low. It is defined as:

$$A_S(\mathbf{p}) = 1 - H\left(\left|\sum_{\mathbf{c_r}\in\mathcal{N}_i(\mathbf{p})} (-1)^{s^{l_p}} V_{ucy}(l_{c_r}^{opt}, l_p)\right| - tsh\right) \tag{4.36}$$

Where N is the number of nodes in $\mathcal{N}(\mathbf{p})$, tsh is a constant parameter that evaluates the strength of the penalty and H(.) is the Heaviside step function. The neighborhood system is designed so that a node in $\mathcal{G}_{r+1,max}$ is connected to several nodes in $\mathcal{G}_{r,max}$. As a results, the penalty of a node is low in two cases. The first and most straightforward case is when the segmentation labels of all the node's neighbors are uncertain. The second case occurs when the neighbors have highly reliable but different labels. This can occur if the node is at the tumor's boundary. In that case, the node will be equally penalized for both labels and the penalty will have almost no impact on the labeling. The segmentation activation criterion measures the strength of the penalty imposed on node $\mathbf{p}$. The case where the labels are equally confident is taken into account by subtracting the penalties associated to different labels. A node will be activated if its penalty strength is smaller than a predefined threshold tsh.

Eventually, the complete MRF energy computed at resolution r and iteration t is written as follows:

$$\begin{aligned} E_c^t(l|V, A\circ\mathcal{T}^{t-1}, \mathcal{G}_r) =& \frac{1}{|\mathcal{G}_r|}\sum_{p\in\mathcal{G}_r} V_p(l_p|V, A\circ\mathcal{T}^{t-1}) \\ &+ \sum_{p\in\mathcal{G}_r}\sum_{q\in\mathcal{N}(p)} V_{pq,}(l_p, l_q) + \sum_{p\in\mathcal{G}_r}\sum_{c\in\mathcal{N}_i(\mathbf{p})} \frac{1}{N} V_{ucy}(l_c^{opt}, l_p) \end{aligned} \tag{4.37}$$

where N is the number of nodes in $\mathcal{G}_{r-1}$ that are connected to node $\mathbf{p}\in\mathcal{G}_r$.

To summarize, at each iteration t, a new grid $\mathcal{G}_r$ and energy are defined based on the deformation field and min-marginals estimated at iteration $t-1$. The first two terms of the energy correspond to the MRF unary and pairwise term. The third term is an inter-level pairwise potential that propagates the previous level's segmentation decisions based on their confidence determined according to their min-marginals. The energy is then minimized and the min-marginals are computed for the next iteration $t+1$.

4.5 Experimental Validation

4.5.1 Data-Set Preprocessing

The healthy reference subject was a 3D FLAIR volume of size $256 \times 256 \times 24$ and resolution $0.9 \times 0.9 \times 5.5mm^3$. The method was tested on a data-set of 110 3D FLAIR volumes of patients featuring DLGGs. All tumors have been manually segmented by experts. The preprocessing step follows the methodology of chapter 3 by regularizing the volumes intensities to the same median and interquartile range as the reference pose. An initial rigid registration of all volumes to the reference pose is performed using the block matching method described in section 3.2.1.1. Eventually, all volumes where skullstripped prior to the registration.

The data-set was split in a training set of 36 volumes randomly selected and a testing set of 74 volumes. The tumor likelihood probabilities are obtained by boosting learning on the 36 volumes. The same features as the ones used in chapter 3 are used, that is, intensity patches and statistics, Gabor features and brain symmetry. The rigid registration step provides a common approximate symmetry plane for all volumes based on the reference pose. To evaluate the potential of our method, the joint framework was compared to a sequential approach. The tumor is first segmented using a pairwise MAP-MRF model, where the likelihood corresponds to the boosting probabilities and an Ising model is used as pairwise cost. Second, a discrete MRF deformable registration is performed by masking the pathology using the obtained segmentation map. Additionally, the potential of the uncertainty driven method is tested by comparing the segmentation and registration results to the ones obtained without uncertainties using a user defined segmentation propagation penalty cost. The results obtained with a maximum resolution and a uniform grid of the same final resolution as the adaptively sampled grid's final resolution are considered.

4.5.2 Implementation

All methods were compared using the set of parameters that yielded the best results. The coarse to fine hierarchical approach consists of 4 grid levels and 3 image levels. The image resolution increases as the grid resolution increases via a Gaussian pyramid, and remains constant at the fourth grid level (see Fig. [4.6]. The resolution increases from $9 \times 9 \times 5$ to $65 \times 65 \times 37$ (maximum). B-splines were used as a projection function, while the inverse projection $\hat{\eta}$ was defined as a linear one as suggested in [Glocker 2008b] to limit the impact of the approximation of the unary term. Because of the interpolation, the obtained segmentation field is not binary and was thresholded

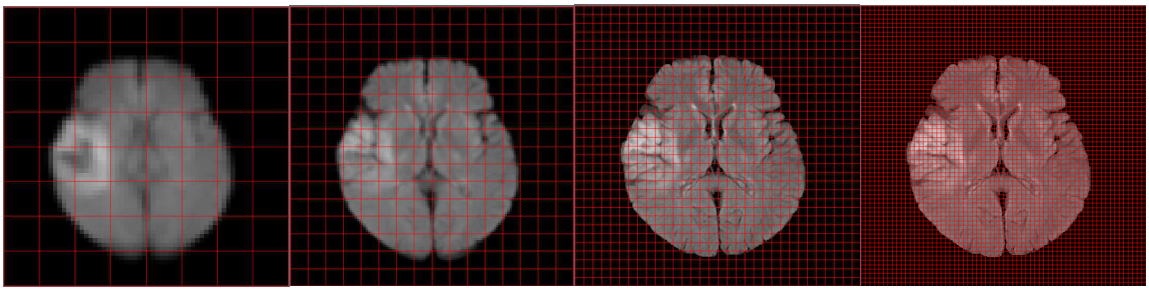


Figure 4.6: *The 4 levels of the incremental approach.*

at 0.5. The registration similarity criterion was the Sum of Absolute Differences. The parameter α is initially set to 1 and is progressively diminished to 0.015. The idea is to give an increasing importance of the presence of the tumor on the registration to avoid an initial bias, while the focus is on segmentation at the highest resolution level that is not necessary for increased registration quality. Only the segmentation labels are considered at this level. The constant cost for registration increases with the grid and image resolution from 5 and 6 times the mean similarity cost value with and without uncertainties respectively. The parameter λ describing the influence of the registration regularization was set to 20 and relaxed in the tumor area to allow for substantial deformations due to the tumor presence. Lastly, the threshold tsh for node activation was set to 1.6, corresponding to a mean uncertainty measure of 0.2 (if neighboring labels have the same value). The same parameters where used for all volumes.

At each grid level, 3 iterations were performed. If the min-marginals are not computed, the displacement sampling is sparse and amounts to 31 labels sampled along the main axes. The sampling is refined at each iteration by reducing the maximum displacement. A dense sampling is adopted to compute the uncertainties at the first iteration, which correspond to 1331 labels in 3D uniformly sampled. Once computed, a sparse sampling is used for the last two iterations. The labels are sampled along the min marginals covariance matrix main axes. The registration uncertainties are not computed at the last two grid levels when α is low and the anisotropy cannot be captured. At these levels, a sparse sampling is used for all iterations. To prevent inadequate displacements towards the tumor area, a small pairwise penalty cost is added to the corresponding labels of nodes at the tumor boundary. The two sampling strategies can be visualized in Fig.[4.7] Given the boosting score likelihoods, the selected parameters yield a run time of about 40 seconds with uniform and sparse sampling and about 3-4 minutes with dense/sparse sampling.

4.5.3 Results

In order to evaluate the use of uncertainties for adaptive sampling, we computed the percentage of activated nodes with respect to the maximal resolution. The mean percentage over all nodes for all levels can be seen in Fig. [4.8] and we can see that a significant portion of nodes is inactive. The

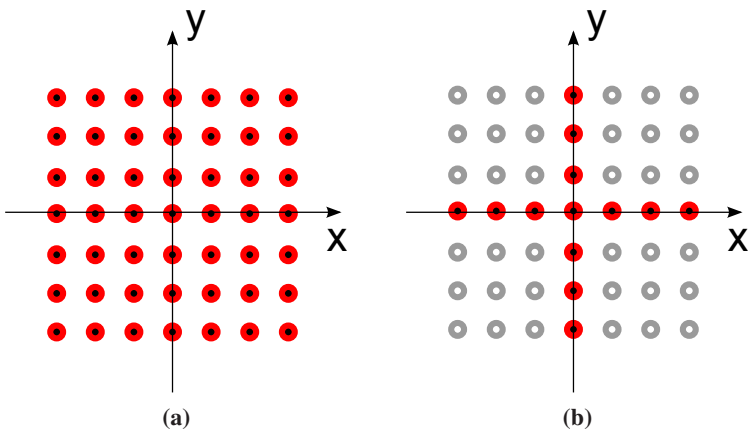


Figure 4.7: *The two displacement sampling strategies, centered around the node's initial position: (a) Dense displacement sampling, (b) Sparse displacement sampling. The red dots correspond to the possible displacements.*

complexity of the framework is considerably reduced, notably at the last, most computationally expensive levels where only 32% and 20 % of the nodes are activated. The relevance of the selected nodes is evaluated visually and several examples can be seen in Fig. [4.9] for different resolution levels. Nodes are activated around the brain's structures and the tumor, where the registration and segmentation have to be precise. This demonstrates the adequacy of the registration and segmentation activation terms. The impact on the run time and memory consumption can be evaluated through a complexity analysis (excluding the time required to build the data term potentials). We consider an MRF with L, E and N being respectively the number of labels, edges and nodes.

- Computational cost: $\mathcal{O}\left(L \times \mathcal{O}\left(E \times N^2\right)\right) \sim \mathcal{O}\left(L \times E \times N^2\right)$ per iteration
- Memory cost: $\mathcal{O}\left(L \times \mathcal{O}\left(N + E\right)\right)$

Reducing the number of nodes to approximately 20% at the finer resolution scale leads to:

- Computational cost: $\mathcal{O}\left(L \times \mathcal{O}\left(0.2E \times (0.2N)^2\right)\right) \sim \mathcal{O}\left(0.008L \times E \times N^2\right)$. When taking into account the number of iterations, we can obtain a complexity that is approximately 3-4 orders of magnitude lower.
- Memory cost: $\mathcal{O}\left(L \times \mathcal{O}\left(0.2(N + E)\right)\right)$, approximately one order of magnitude lower.

The quality of the segmentation was evaluated using the Dice Score, False Positive Rate (FP), True Positive Rate (TP) and Mean Absolute Distance between contours (MAD) with respect to the manual segmentation. We compared the segmentations obtained by simple MAP-MRF segmentation, concurrent segmentation and registration at high and low resolution, and uncertainty driven concurrent segmentation and registration. The joint framework shows increased performance in

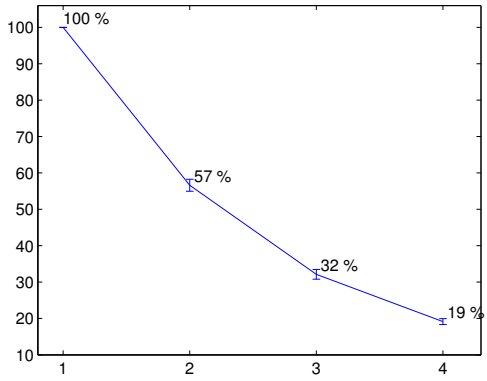


Figure 4.8: *Mean percentage of activated nodes per resolution level.*

the segmentation of tumors, notably reducing the number of false positives which is highlighted by the lowest FP rate and MAD score (false detections outside the tumor region can strongly impact the MAD score). Equivalent results are obtained with high resolution uniform sampling and adaptive sampling, while segmentations of poor quality are obtained with the low resolution uniform sampling. The quantitative results can be seen in Fig.[A.4], while visual results are shown in Fig.[4.13].

The registration quality is evaluated mostly qualitatively. The individual registration method with masked pathology (MAP-MRF segmentation) is compared to the joint framework without uncertainties at different resolutions as well as with the graphical model endowed with uncertainties. For quantitative evaluation, the ventricles where segmented on 33 volumes where the tumor was close to the ventricles. The Dice and MAD scores, FP and TP rates were computed between the segmented and registered ventricles using the different methods outside the tumor area. Quantitative results show equivalent performance between the individual and joint framework. The performance is preserved with the adaptive sampling while the quality diminishes when an equivalent (in term of nodes) uniform low resolution sampling. The quantitative results are shown in Fig.[4.11], while qualitative results are shown in Fig.[4.12]. We observe higher performance around the tumor area where the individual framework can fail.

4.6 Discussion

This chapter introduced a concurrent tumor segmentation and registration framework that directly exploits the dependencies between both problems. We adopt a discrete Markov Random field model, where the tumor is detected by boosting classification as well as a segmentation through registration term that detect areas with high dissimilarities. The registration similarity criterion is relaxed in the tumor area and as a byproduct of that, this area is deformed by interpolation with

the displacement at the boundary. The principal shortcomings of discrete methods are addressed via the evaluation of the local segmentation and registration uncertainties and their exploitation for adaptive displacement set and sparse grid sampling. Segmentation results show a notable improvement towards eliminating false positives and an increased robustness due to the hierarchical approach, while the registration is smoother and preserves the deformed image's anatomy in the tumor area. The content driven grid sampling strongly reduces the complexity of the model with a maintained performance.

The choice of pathology masking for registration (justified by the infiltrative nature of DLGGs), along with the discrete formulation, results in a fast and versatile algorithm. Indeed, the method is modular with respect to the registration similarity criterion and the tumor detection method. No assumption is made on the pathology nor its progression, which makes the method easily adaptable to other pathologies and clinical contexts where correspondences are missing. It should be noted that in the context of fast growing, space occupying lesions, integrating tumor growth information would be necessary. Such tumors can impact the brain structures where a relaxation of the pairwise cost at the tumor boundary wouldn't be sufficient.

Several improvements of the algorithm could be considered. Our current implementation of the adaptive sampling simply discards the inactive nodes in the evaluation of the optimal labeling and associates them with infinite costs to changes of labels. This evaluates the impact of reducing the grid's complexity on the quality of the results but does not yet impact the run time. A direct construction of the non-uniform grid would strongly decrease the run time (both in terms of optimization and data term construction) and increase the robustness and quality of the results as active nodes would have direct edges connecting them.

The tumor segmentation results remain dependent on the initial boosting detection and partial segmentations will be obtained if the tumor is poorly detected. A first step in order to obtain better segmentation would be to focus on the detection step. The two most straightforward ways of improving the quality of the detection are the addition of multimodal information such as T1 and T2 volumes that can be easily introduced in the classification feature vector, and assigning the different parts of the tumor (notably necrosis and active tumor) to different classes. The discrete labeling formulation makes the extension to multi-class segmentation straightforward.

The algorithm presented is modular with respect to the registration similarity criterion, but optimal interdependency between segmentation and registration can only be obtained using intensity based similarity measures, that can roughly detect the tumor area via the sole registration unary term. The algorithm is therefore better suited to monomodal registration. An extension of the method to multimodal registration in the context of tumor resection is presented in appendix A.

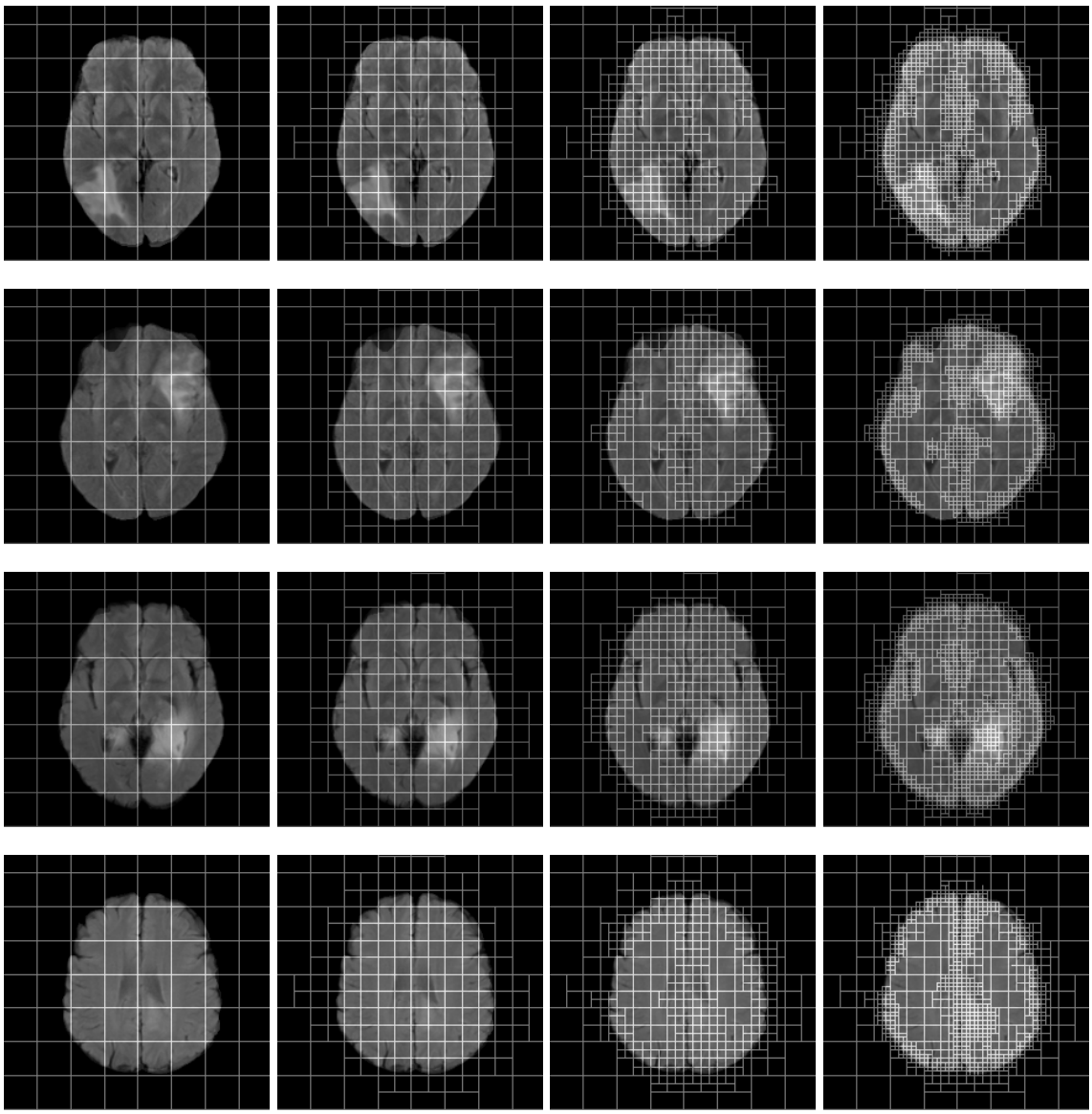


Figure 4.9: *Visual examples of the activated nodes for the 4 levels of the incremental approach. The nodes are superimposed to the target image.*

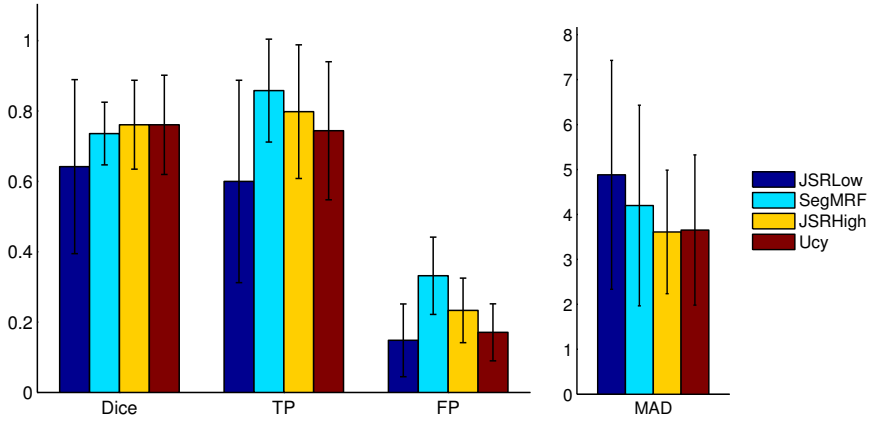


Figure 4.10: *Quantitative Segmentation Results: Error bars (mean and standard deviation) of the Dice score, False Positive (FP) and True Positive (TP) rates and MAD score for the joint framework with low (JSRLow) and high resolution (JSRHigh), the MAP-MRF segmentation framework (SegMRF) and the uncertainty based approach (Ucy).*

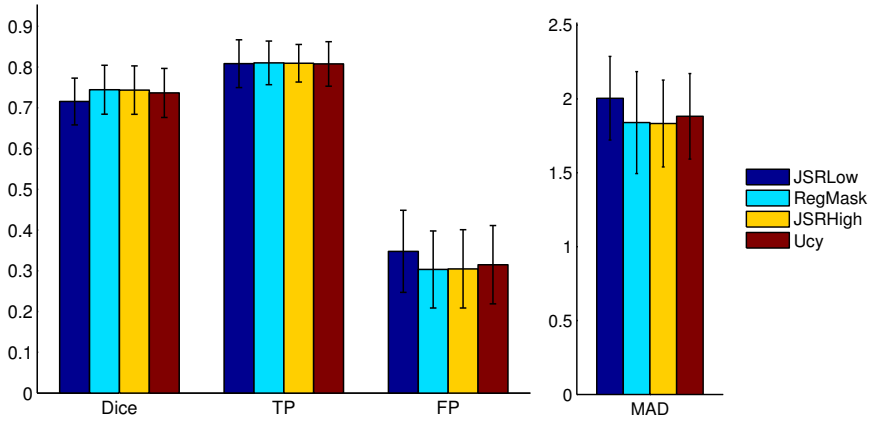


Figure 4.11: *Quantitative Registration Results: Error bar graphs of the Dice, True Positives (TP), False Positives (TP) and MAD scores obtained for the joint framework with low (JSRLow) and high resolution (JSRHigh), the individual registration framework with masked pathology (RegMask) and the uncertainty based approach (Ucy).*

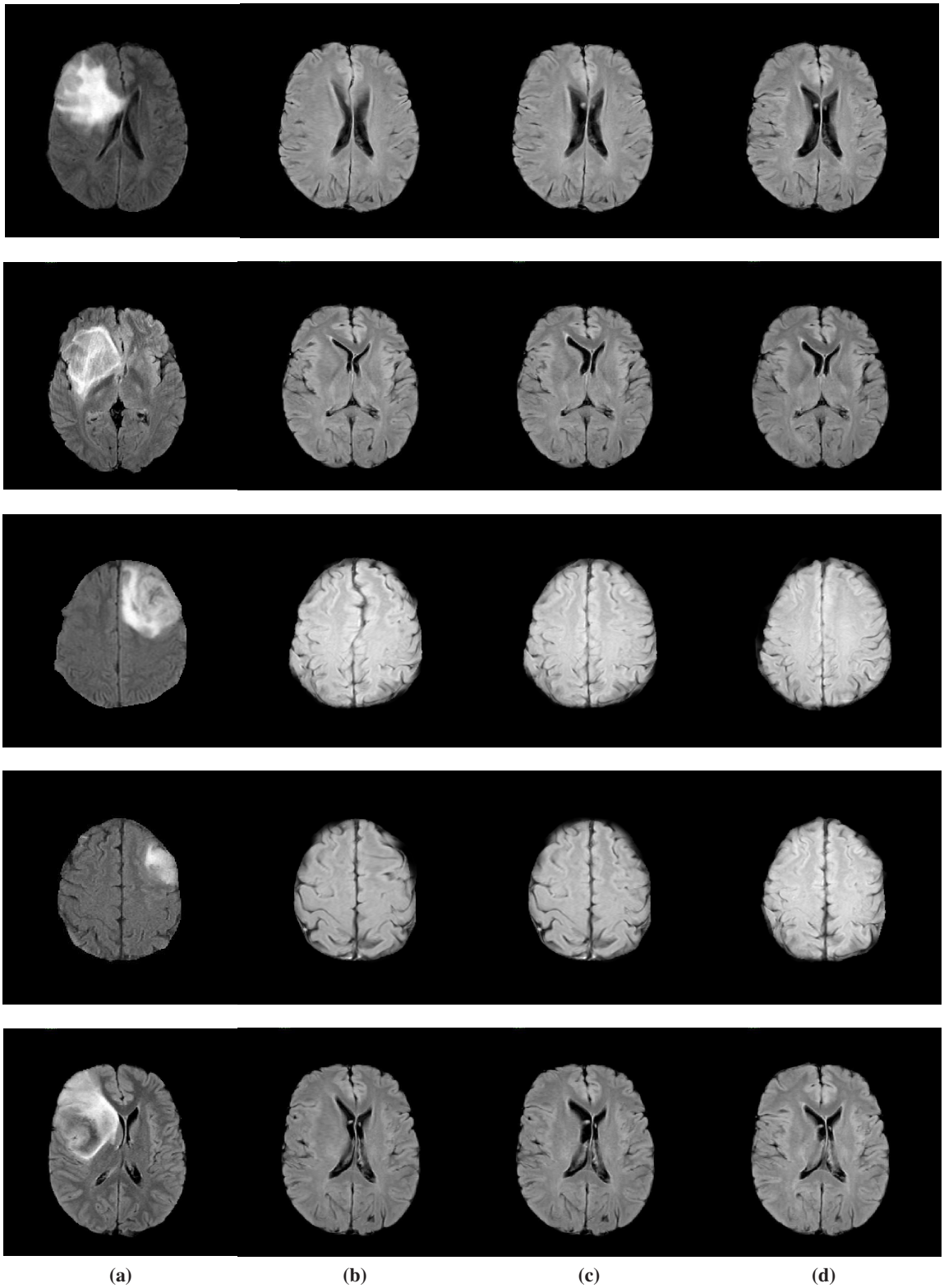


Figure 4.12: *Visual examples of Registration results. (a) Target image, deformed image obtained using (b) the individual framework with masked pathology, (c) the Joint framework, high resolution, (d) the Joint framework, with adaptive sampling.*

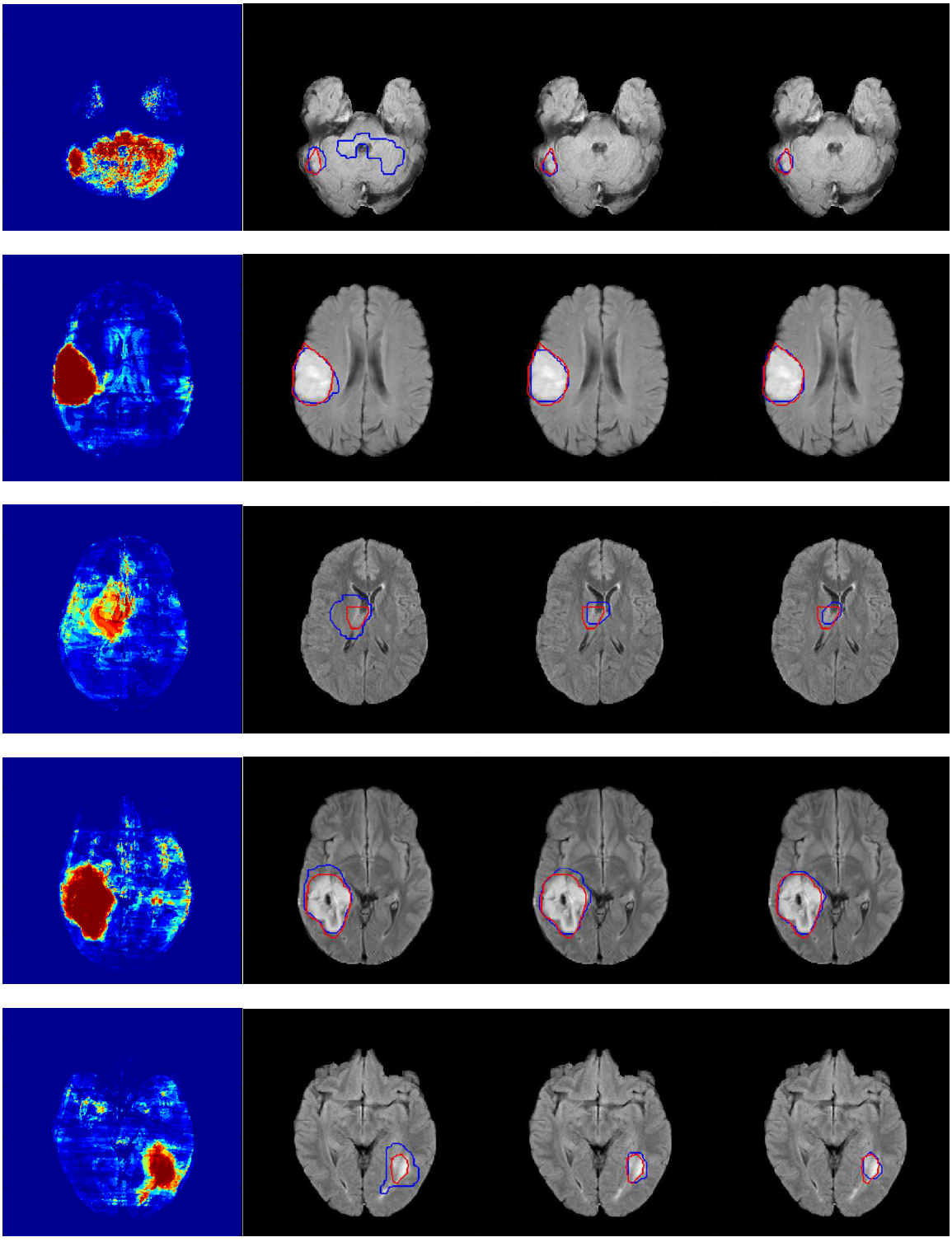


Figure 4.13: *Visual examples of Segmentation results. (a) Boosting Score, (b) MAP-MRF framework, (c) Joint framework, high resolution, (d) Joint framework, with adaptive sampling. Automatic segmentations (blue) are compared to the manual segmentation (red)*

Chapter 5

Conclusion

The main objective of this thesis was the development of methods for tumor segmentation through atlas based prior knowledge. Two approaches have been proposed: one that exploits a tumor specific atlas as a spatial prior, and a second exploiting anatomical knowledge. The presented contributions are not limited to tumor segmentation as we proposed tools for the study of diffuse low-grade gliomas, registration with missing correspondences, concurrent registration and segmentation, and adaptive sampling for hierarchical grid based segmentation and registration.

5.1 Contributions

The first contribution of this thesis is the construction of a probabilistic atlas from 210 different patients mapping the preferential locations of DLGGs. This is carried out through a graph structure encoding the spatial proximity between tumors over 210 affinely registered volumes. The preferential locations are obtained by regrouping nearby tumors in clusters. This yields an eleven-components graph where each cluster corresponds to a preferential location and potential position dependent behavior. This atlas paves the way for studies on the genesis and natural history of the tumors. Understanding the tumors better is key towards optimizing their treatment for a given patient.

A direct application of the atlas was proposed for tumor segmentation and characterization via a coupled Markov Random Field model. The method simultaneously assigns a new tumor to one of the atlas' clusters and labels the tumor voxels. The atlas plays the role of a prior on the tumor voxels' localization based on the spatial repartition of tumors voxels in the cluster. This can be seen as the association of local and global information where determining the tumor's global category (cluster) allows a more refined local segmentation based on the knowledge of its position and expected spatial extension directions. It is clear that this method is not limited to atlas based low-grade glioma segmentation but can be considered in any segmentation process where the objects to be segmented tend to appear in predefined locations or follow a repetitive structural pattern.

The second important contribution of the thesis is the design of a concurrent segmentation and

registration method with missing correspondences. The method is among the first to tackle the problem of simultaneous registration and segmentation when the object to be segmented is absent from one of the images. The deformation and segmentation fields are estimated concurrently on a sparse grid using an MRF model where both solution spaces are interconnected. The proposed method is characterized by its modularity and scalability. It is indeed, designed in a very general way. Almost no assumptions are made on the pathology and on the object detection method, as well as the registration criterion. This modularity is notably highlighted in appendix A where a direct application of the concurrent registration and segmentation method is proposed in the context of surgical resection. A strong asset of the method is its speed, which justifies its extension to surgical interventions.

In the context of concurrent registration and segmentation, we proposed a technique in order to address one of the most important shortcomings of discrete and hierarchical registration and segmentation methods that is the trade off between precision and complexity. We proposed an adaptive sampling approach based on local uncertainties. The precision (higher resolution) is determined based on the reliability of the evaluated segmentation/registration labels. This leads to a considerable decrease of the complexity at high resolutions with a focus on areas with discontinuities and tumors boundaries. Furthermore, it provides a qualitative manner for the interpretation of the obtained results. The combined registration and segmentation method is designed in a way that allows its application to the sole problem of registration or segmentation.

5.2 Future Work

The most straightforward future development is the association of the two proposed segmentation methods in one single framework. Using manual segmentations as detection probabilities, the concurrent segmentation and registration (CSR) framework could be used for the construction of a deformable low-grade glioma atlas on the preferential locations of the tumors. The use of deformable registration would increase the precision as well as allow studies on the impact of the tumors on the brain structures and functional areas. This deformable atlas could then be incorporated as a spatial prior in the CSR method, while cluster assignment could be based on initial detections in the hierarchical framework.

Throughout this thesis, a single subject anatomical template has been used as a reference pose. This introduces a bias in the construction of the atlas as well as increases the difficulty of the registration problem when the patient and reference have strong anatomical differences. The construction and use of an averaged population template of the FLAIR modality would be more accurate and adapted to the latter problem. As it encompasses the inter subject variability, deformations would be more natural. Such an atlas would also reduce the bias on the former problem. Furthermore, atlas construction has raised a lot of interest as there has been many attempts at eliminating this bias [Zollei 2005, Bhatia 2004, Joshi 2004]. A notable approach for MRF based registration [Sotiras 2009] could be considered in our context.

The use of growth models [Jbabdi 2005, Konukoglu 2010, Rekik 2013] could be considered in

order to increase the quality of both the atlas construction and CSR framework. Indeed, growth models would enable a more precise localization of the origin point of the tumor which could differ from the center of gravity [Rekik 2013], while including growth models in the registration/segmentation framework could prove useful in the context of heavily deformed brains.

Multimodality would most likely bring useful additional information for both segmentation and registration. The different tumor appearances on multimodal images would make the tumor detection step more accurate and consequently increase the quality of the overall segmentation results. Diffusion Tensor Imaging is a very promising modality that could be relied upon for registration and segmentation. While providing structural information for increased robustness and anatomical structures preservation, it can constitute a strong cue on the tumor's (or resection cavity) position as the tumor tends to displace or disrupt the fibers' natural organization. Integration of this knowledge in the CSR framework could have a strong positive impact on the results.

Last but not least, the use of recent parallelized message passing optimization techniques [Alchatzidis 2011] would further speed up the methods towards real time as well as provide min-marginals computation in a single optimization step.

Publications of the Author

International Journals

- Sarah Parisot, William M. Wells III, Stéphane Chemouny, Hugues Duffau and Nikos Paragios. Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. Medical Image Analysis (in press).

International Conferences

- Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Graph Based Spatial Position Mapping of Low-Grade Gliomas*. In G. Fichtinger, A.Martel, and T. Peters: MICCAI, Part II, LNCS 6892, pp. 508-515. Springer, Heidelberg, 2011.

- Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Graph-based detection, segmentation & characterization of brain tumors*. In IEEE Conference on Computer Vision and Pattern Recognition - CVPR, pages 988–995. IEEE, 2012.

- Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Joint tumor segmentation and dense deformable registration of brain MR images*. In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editeurs, MICCAI (2), volume 7511, pages 651–658. Springer, 2012. **Oral presentation.**

- Sarah Parisot, William Wells III, Stephane Chemouny, Hugues Duffau and Nikos Paragios. *Uncertainty-Driven Efficiently-Sampled Sparse Graphical Models for Concurrent Tumor Segmentation and Atlas Registration*. In The IEEE International Conference on Computer Vision (ICCV), December 2013.

Appendix A

Concurrent Segmentation and Registration Application: Registration of Preoperative and Intraoperative Images with Tumor Resection

A.1 Introduction

The treatment of DLGGs has long been controversial. Based on several retrospective studies, the positive impact of surgical resection on the patient's progression free and overall survival is undeniable and maximal safe tumor resection is now considered as the first therapeutic option for DLGGs. As a result, the surgical resection task is very frequent and an essential part of the treatment of the tumor. The process is very challenging as the surgeon seeks to remove as much tumor as possible while preserving the critical tissue associated to functional areas.

Medical images are heavily used for surgery planning. Structural imaging enables to identify the position and extent of the tumor, while functional images allows mapping of the functional areas with respect to the tumor. Such data allows to estimate the volume of possible resection. During surgery, the preoperative data is often used in order to guide the surgeon and provide additional information on the brain's functional and structural organization. However, the preoperative image is generally aligned using rigid or affine registration which could be highly unreliable due to important deformations of the brain during surgery (mainly due to brain shift and tissue collapse). Those deformations are due to the resection as well as a brain shift during surgery caused notably by gravity, tumor mass effet, edema and CSF leakage [Hartkens 2003].

The development of intra operative MR imaging enabled to obtain anatomical data and increased the efficiency of the procedure through reliable anatomical guiding. However, surgical planning (mainly, the expected amount of tumor resection) as well as additional functional and structural information (fMRI, DTI for instance) acquired prior to the surgery. Such modalities

provide important information for neurosurgical guidance. Direct acquisition of such data during surgery is currently not an option due to long acquisition and processing times, and the presence of strong distortions.

For those reasons, aligning the preoperative data to the intraoperative image via non-rigid registration is highly desirable, while the segmentation of the resected area would enable direct evaluation of a potential residual tumor. Similarly to the lacking correspondences induced by the tumor with respect to a healthy subject, the registration task is here violated by the presence of a resected area that is not part of the preoperative image. Specific methods have to be designed in order to accommodate for the presence of this cavity.

Chapter 4 introduced a concurrent tumor segmentation and registration method in the presence of DLGGs that is easily adaptable to other clinical contexts. This chapter presents an extension of the method for the context of surgical brain tumor resection. We propose a concurrent preoperative to intraoperative registration method that is coupled with a segmentation of the resection cavity. A probabilistic formulation is proposed, where the registration task relies on the intensity joint histogram and the resected area is detected based on its intensity profile as a mixture of Gaussians. The knowledge on the position of the tumor in the preoperative image is exploited as a spatial prior on the position of the resection cavity.

The remainder of this chapter is organized as follows: section A.2 presents an overview of the non-rigid registration methods in the presence of a tumor resection, including methods that are coupled with a segmentation of the resection area. Section A.3 presents a MAP-MRF probabilistic interpretation of the concurrent segmentation and registration method while section A.4 describes the definition of the segmentation and matching terms. Experimental validation is part of section A.5 and Discussion concludes the chapter.

A.2 Related work

This section presents prior work on non-rigid registration with tumor resection. The general case of pathology masking presented in section 4.2 could also be applied in this context. Aside from the related work on registration with missing correspondences, there has been little focus on tumor resection and most specifically preoperative to intraoperative non-rigid registration.

Clatz et al. [Clatz 2005] presented a block matching method coupled with a biomechanical modeling of the brain. The block matching method, introduced in chapter 3, seeks correspondences between images blocks by minimizing local similarity measures. This results in a sparse estimation of the deformation field that is in a second step estimated on the whole volume. With the objective to have as little computational overhead as possible during surgery, several preprocessing steps are carried out based on the preoperative image. The brain, ventricles and tumor are segmented on the preoperative image and those segmentations are used to build a biomechanical model of the brain based on the Finite Elements model. In order to avoid comparing homogeneous and ambiguous blocks, a block rejection step is also carried out preoperatively, by measuring the block intensity variance and selecting only the blocks with the highest variance. A prohibited

connectivity criterion is included to prevent selected blocks to have overlapping structures. Eventually, blocks are selected outside the tumor area to prevent matchings with the resection cavity. Last but not least, they construct a structure tensor that measures the smoothness of the intensity distribution and provide a measure of confidence of the block's displacement. During surgery, the matchings between blocks are estimated using the Cross Correlation. The dense deformation field is then estimated using the biomechanical model. The influence of a block on the deformation field depends on the value of the similarity metric and the direction of the displacement with respect to the tensor structure (the matching should be colinear with the intensity gradient direction in the block). Similarly to the method of [Ourselin 2000], an additional block rejection step using least trimmed squares is included to reject resection blocks. The method was tested "offline" on 6 pre and intraoperative T1 images, and more thoroughly in [Archip 2007] on 11 patients during surgery. In [Archip 2007], the data was also fused with fMRI and DTI data.

Risholm et al. [Risholm 2009, Risholm 2010] proposed a registration framework based on the bijective demons algorithm. They adopt a fluid anisotropic regularization, where the amount of regularization differs from one direction to the next. The resected area was detected using level-sets. The speed function α defining the evolution of the front along its normal direction is defined based on the intensity disagreement between the source and target image:

$$\alpha(\mathbf{x}) = |J(\mathbf{x}) - (I \circ \mathcal{T})(\mathbf{x})| e^{-\frac{|\nabla I|}{\sigma}} - L \tag{A.1}$$

I and J are the preoperative and intraoperative images respectively, $\mathcal{T}$ is the transformation and L is a threshold parameter. Given this definition, the front expands when the intensity disagreement is strong and retracts when it is small, based on threshdold L, while the exponential term's role is to eliminate the propagation of the front at large gradients. The estimated resection area is exploited for registration as a diffusion sink. The idea is that instead of propagating the computed demon velocities over the whole image, the forces computed inside the resection area (that are not reliable) do not propagate beyond the boundary. In other words, diffusion is allowed towards the resection cavity's boundary, but not from it. This diffusion sink is simply defined by setting the anisotropic regularization parameters to 0 in the appropriate direction. The algorithm alternates between performing the registration given the segmented resected area and diffusion sink constraint and evaluating the resection area. Experiments were performed on 2D data in [Risholm 2009] and 3D in [Risholm 2010].

An Expectation-Maximization framework was introduced in [Liu 2010]. They proposed a point based registration framework inspired from the method of [Chui 2003]. Given an initial set of points extracted from the source and target image (canny edge detection) and an estimation of the resected area M^t, the method estimates the correspondences between the points as a normal distribution on the intensity difference between both sets of points. It also registers the images by minimizing the distance between the matched points (E-step). The points within the resected area are not taken into account (*i.e.* they are removed from the set of points in the source and target image). The deformation field on the whole image is recovered through biomechanical modeling and the Finite Element method. The M-step reestimates the resected area using the deformed mesh

and a segmentation of the brain. Elements in the mesh that are contained in the background and constitute the largest connected mesh with the currently estimated resection mesh M^t are added to the updated resection mesh M^{t+1}. Points of the source image that belong to this mesh are removed from the set.

Chitphakdithai et al. [Chitphakdithai 2010] adopted a Bayesian formulation inspired from the work of [Pohl 2006] for the registration of preoperative and intraoperative images, as well as the segmentation of the brain and the resection cavity. The registration problem is posed as a marginalized MAP estimation of the deformation field:

$$\mathcal{T}^{opt} = \underset{\mathcal{T}}{\operatorname{argmax}} \log \sum_{\mathcal{S}} p(\mathcal{T}, \mathcal{S}|I, J) \tag{A.2}$$

The problem is solved using the EM algorithm, where iteratively, the E-step estimates the segmentation map $\mathcal{S}$ given the current estimation of the transformation $\mathcal{T}^t$ and the M-step updates the transformation $\mathcal{T}^{t+1}$ given the estimated segmentation map. The M-step is written as follows:

$$\begin{aligned} \mathcal{T}^{t+1} = \underset{\mathcal{T}}{\operatorname{argmax}} \sum_{\mathbf{x}} \sum_{l} \\ \log p(\mathcal{S} = l|I, J, \mathcal{T}^k) \left(\log p(I(\mathcal{T}(\mathbf{x}))|J, \mathcal{S} = l, T) + \log p(J(\mathbf{x})|\mathcal{S}(\mathbf{x}) = l)\right) + \log p(\mathcal{T}) \end{aligned} \tag{A.3}$$

where the label "l" can correspond to three classes: background, brain or resection. The term $p(I(\mathcal{T}(\mathbf{x}))|J, \mathcal{S} = l)$ corresponds to the similarity criterion for registration. In areas with correspondences (background and brain), it is modeled as a normal distribution with respect to the intensity difference between both images, with a class specific standard deviation. It is set as a constant in the resection area. The term $p(J(\mathbf{x})|\mathcal{S}(\mathbf{x}) = l)$ is an intensity-based prior on the intraoperative image's intensity given a class. It is set as a uniform distribution in the classes that have correspondences and defined as a Maximum Likelihood estimate in the resection area based on a set of training data. Last, the transformation prior $p(\mathcal{T})$ is a smoothness term acting on the deformation field, expressed through Free Form Deformations. The E-step computes the segmentation posterior probabilities. The method was tested in synthetic data and 6 pairs of 3D clinical data using Leave One Out cross validation for resection probability learning. The method was extended in [Chitphakdithai 2011] by the introduction of an MRF smoothing prior on the segmentation field.

A.3 Probabilistic Formulation

This section proposes a probabilistic adaptation of the concurrent segmentation and registration MRF framework. Let us consider the source preoperative image I of a subject with a tumor defined on a domain Ω, and the intraoperative target image J of the same subject during the resection of the tumor. We assume that the tumor in the preoperative image has been segmented prior to the surgery where $\mathcal{S}_\mathcal{I}$ is the segmentation map. We seek the transformation $\mathcal{T}$ that maps the preoperative image to the intraoperative image as well as the resection cavity segmentation map $\mathcal{R}$. A sparse grid $\mathcal{G}$ is superimposed to the source volume I, and each grid node $\mathbf{p}$ is to be assigned a label

$l_p \in \mathcal{L} = \{1, .., 2n\}$. Each label corresponds to a pair segmentation/displacement $\{\mathbf{d}^{\mathbf{l_p}}, s^{l_p}\}$ where the displacements are selected among a predefined discrete set $\Delta = \{\mathbf{d^1}, ..., \mathbf{d^n}\}$. Once the grid nodes' labels are obtained, the transformation and segmentation fields are computed on the whole image by interpolation:

$$\begin{aligned} \mathcal{T}(\mathbf{x}) &= \mathbf{x} + \sum_{\mathbf{p}} \eta(\|\mathbf{x} - \mathbf{p}\|)\mathbf{d}^{\mathbf{l_p}} \\ \mathcal{R}(\mathbf{x}) &= \sum_{\mathbf{p}} \eta(\|\mathbf{x} - \mathbf{p}\|)\mathbf{s}^{\mathbf{l_p}} \end{aligned} \tag{A.4}$$

Where $\eta(.)$ is the projection function.

The problem can be cast as a MAP-MRF estimation of the posterior probability $P(\mathbf{l}|I, J)$ on the labeling $\mathbf{l} = \{l_p\}_{\mathbf{p}\in\mathcal{V}}$ given the observed images :

$$\hat{\mathbf{l}} = \underset{\mathbf{l}\in\mathcal{L}}{\operatorname{argmax}}\, P(\mathbf{l}|I, J) \propto P(I, J|\mathbf{l})P(\mathbf{l}) \tag{A.5}$$

The prior probability $P(\mathbf{l})$ is modeled as a pairwise MRF defined on the sparse grid $\mathcal{G}$:

$$P(\mathbf{l}) = \exp\left(-\left(\sum_{\mathbf{p}\in\mathcal{G}} V_p(l_p) + \sum_{\mathbf{p}\in\mathcal{G}}\sum_{\mathbf{q}\in\mathcal{N}(\mathbf{p})} V_p, q(l_p, l_q)\right)\right) \tag{A.6}$$

Following the methodology of chapter 4, the pairwise term is defined as the sum of a registration and segmentation smoothing prior weighted by the distance between the nodes:

$$V_{p,q}(l_p, l_q) = \lambda_s \frac{|s^{l_p} - s^{l_q}|}{\|\mathbf{p} - \mathbf{q}\|} + \lambda_r \frac{\left(\mathbf{d}^{\mathbf{l_p}} - \mathbf{d}^{\mathbf{l_q}}\right)^2}{\|\mathbf{p} - \mathbf{q}\|} \tag{A.7}$$

The unary term $V_p(l_p)$ is a spatial prior that exploits the fact that the localization of the tumor in the preoperative image gives insight on where the resection cavity is going to be. Indeed, the resection area is expected to be in the preoperative tumor area or within its close vicinity. The preoperative tumor segmentation is used to define the area where the resection cavity is expected to be and not to be. To allow segmentation outside the tumor area, the segmentation map is smoothed by Gaussian filtering.

$$V_p(l_p) = \begin{cases} -\int_\Omega \bar{\eta}(\|\mathbf{x} - p\|) \log \mathcal{S}_{I,s}(\mathbf{x} + d^{l_p})dx & \text{if } s^{l_p} = 1 \\ -\int_\Omega \bar{\eta}(\|\mathbf{x} - p\|) \log(1 - \mathcal{S}_{I,s}(\mathbf{x} + d^{l_p}))dx & \text{otherwise} \end{cases} \tag{A.8}$$

where $\mathcal{S}_{I,s}$ is the smoothed tumor segmentation map, and $\bar{\eta}(.)$ is the back projection function that describes the influence of voxel $\mathbf{x}$ on control point $\mathbf{p}$. We recall that the transformation $\mathcal{T}(\mathbf{x})$ is approximated in the unary terms as $\mathbf{x} + d^{l_p}$ in order to preserve the independency asumption inherent to MRF models.

Assuming independency between the observations and the projection of the likelihood probability on the grid control points leads to a MAP estimation problem that aims at minimizing the following objective function, defined on the sparse grid $\mathcal{G}$:

$$E(\mathbf{l}) = \sum_{p \in G} \left(V_p(l_p) - \int_\Omega \bar{\eta}(\|\mathbf{x} - p\|) \log P(I(\mathbf{x}), J(\mathbf{x})|l_p) dx \right) + \sum_{\mathbf{p} \in \mathcal{G}} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} V_p, q(l_p, l_q) \quad \text{(A.9)}$$

The obtained MRF energy is minimized using the Fast-PD algorithm.

A.4 Likelihood term

The next step consists in defining the likelihood term $P(I(\mathbf{x}), J(\mathbf{x})|l_p)$ that will play an essential part in the concurrent registration and segmentation framework. Two cases have to be distinguished: if we are in the background ($s^{l_p} = 0$) we want the joint probability to be maximized when the images are aligned while in the resection cavity area ($s^{l_p} = 1$) no correspondences should be sought. The likelihood must enable the detection of the resection cavity as well as the matching of both images.

A.4.1 Resection Cavity Detection

The intensity profile associated to a resection cavity is simpler than the one associated to brain tumors. The detection is therefore easier and simple classification methods can be used in order to separate it from the rest of the image. The resection cavity likelihood term is defined based on the observation that it is made of two possible elements: air and fluid (CSF or blood). When working on T2 images, the air appears as the darkest intensity class while fluid is the brightest. The resection area can be isolated from the rest of the image by modeling the intensity of the area delineated by the spatial prior (that includes the resected area) as a mixture of Gaussians and selecting the ones that correspond to its intensity profile.

A.4.1.1 Gaussian Mixture Models

The Gaussian Mixture Model (GMM) is a popular classification technique that models the intensity distribution as a mixture Gaussian distributions. By choosing a sufficient number of Gaussians, it is possible to capture the intensity distribution of different image classes. The Gaussian mixture distribution is written as follows:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) \quad \text{(A.10)}$$

where K is the number of Gaussians, and (μ_k, Σ_k) are their respective mean and standard deviation. π_k is a mixture coefficient describing the influence of the Gaussian on the overall distribution. The mixture coefficients must satisfy $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$.

The classification can be performed in a supervised or unsupervised setting. In the supervised setting, a class dependent GMM is learned on the intensity distribution over all training samples of the same class. This yields a per class probability dependent on the image intensities. The unsupervised setting learns a GMM on the whole image's intensity distribution and a class can be associated to one or several learned Gaussians. In both cases, the Gaussians parameters are estimated from the set of samples $\mathbf{X} = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$ by maximizing the log likelihood function using the Expectation-Maximization algorithm:

$$\log p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma}) = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x_n}|\mu_k, \Sigma_k) \right) \tag{A.11}$$

A.4.1.2 Expectation Maximization Algorithm

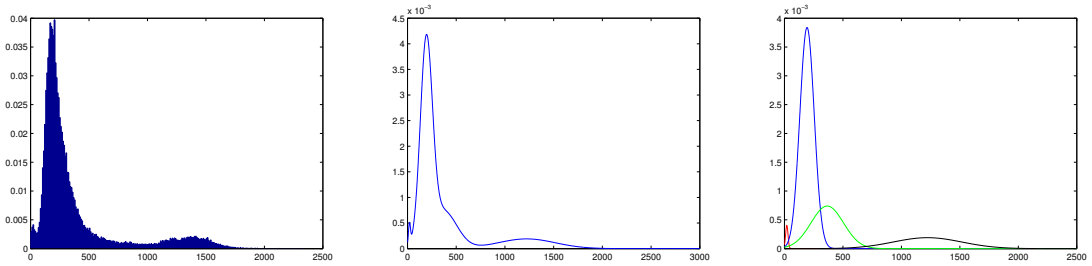
The Expectation-Maximization algorithm [Dempster 1977] is an iterative procedure for maximum likelihood estimates in the presence of missing data. Consider X the observed data and Z the missing data. Their joint distribution $p(X, Z|\theta)$ is governed by a set of parameters θ and its maximization is supposed to be straightforward (complete log likelihood). Since Z is not observed, we need to maximize the incomplete log likelihood function:

$$\log p(X|\theta) = \log \left(\sum_{Z} p(X, Z|\theta) \right) \tag{A.12}$$

Unfortunately, the optimization is not trivial due to the presence of the sum that results in complicated expressions. The problem can be approached by acknowledging that for any distribution $q(Z)$ defined over the missing data, we have:

$$\begin{aligned} \log p(X|\theta) = &\quad \mathcal{L}(q, \theta) + KL(q\|p) \\ \text{where } \mathcal{L}(q, \theta) = &\quad \sum_{Z} q(Z) \log \left(\frac{p(X, Z|\theta)}{q(Z)} \right) \\ KL(q\|p) = &\quad -\sum_{Z} q(Z) \log \left(\frac{p(Z|X, \theta)}{q(Z)} \right) \end{aligned} \tag{A.13}$$

KL is the Kullback-Leibler divergence between distributions $q(Z)$ and $p(Z|X, \theta)$ and is therefore always positive, and null when $q(Z) = p(Z|X, \theta)$. $\mathcal{L}(q, \theta)$ constitutes thus a lower bound on $\log p(X|\theta)$. The EM algorithm aims at maximizing this lower bound and consists of two steps that are repeated iteratively. Given a prior estimation of the parameters $\theta^{(t)}$, the Expectation step (E-step) maximizes the lower bound $\mathcal{L}(q, \theta)$ with respect to $q(Z)$ while holding θ fixed. Following equation A.13, this maximum is obtained when $q(Z) = p(Z|X, \theta^{(t)})$.



(a) **(b)** **(c)**

Figure A.1: *Gaussian Mixture Model of the resection cavity intensity. (a) Intensity histogram, (b) Modeled intensity distribution, (c) Gaussians constituting the GMM.*

The Maximization step (M-step) maximizes $\mathcal{L}(q, \theta)$ with respect to θ given the current estimation of $q(Z)$:

$$\begin{aligned}\mathcal{L}(q,\theta) &= \sum_{Z} p(Z|X,\theta^{(t)}) \log p(X,Z|\theta) - \sum_{Z} p(Z|X,\theta^{(t)}) \log p(X,Z|\theta^{(t)}) \\ &= Q(\theta,\theta^{(t)}) + \text{cst}\end{aligned} \tag{A.14}$$

$Q(\theta, \theta^{(t)})$ is defined as the expectation of the complete log likelihood and is what is maximized during the M-step. In summary, the EM algorithm alternates between computing a lower bound of the log likelihood for the current parameters, and maximizing this lower bound to obtain the new parameter values.

$$\begin{aligned}\text{E-step } q^{(t+1)}(Z) &= p(Z|X,\theta^{(t)}) \\ \text{M-step } \theta^{(t+1)} &= \underset{\theta}{\operatorname{argmax}}\, Q(\theta,\theta^{(t)})\end{aligned} \tag{A.15}$$

A.4.1.3 EM for Gaussian Mixture Models

For Gaussian Mixture Models, we can introduce a K-dimensional binary random variable $\mathbf{z}$ that corresponds to the hidden data [Bishop 2006]. Each observed variable $\mathbf{x_n}$ is associated with a hidden variable $\mathbf{z_n}$ that defines which probabilty distribution the sample follows. Taking into account this variable, the mixture coefficients can be redefined as $\pi_k = p(z_k = 1)$ and can be seen as the prior probability of $z_k = 1$. Estimating the Gaussians parameters amounts to estimating the hidden variable's posterior probability or *responsibility* $\gamma(z_n k)$ during the E-step given the current

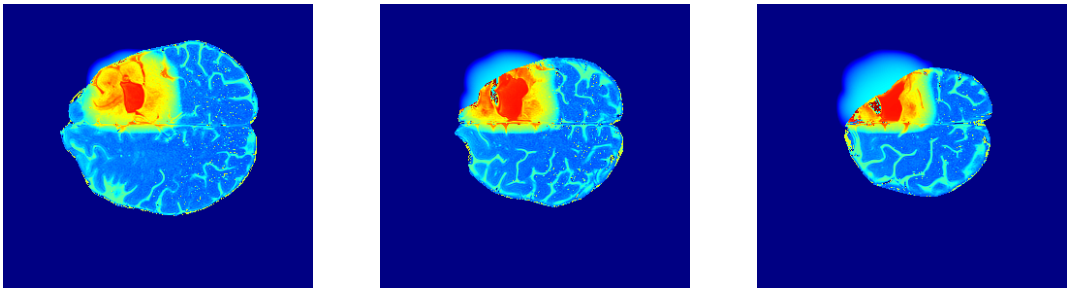


Figure A.2: *Log probabilities associated to the resection cavity coupled with the spatial prior.*

parameters:

$$\begin{aligned} p(z_{nk}=1|\mathbf{x_n},\mu_k^t,\Sigma_k^t,\pi_k^t) &= \gamma(z_{nk}) \\ &= \frac{p(z_{nk}=1)p(\mathbf{x_n}|z_{nk}=1)}{p(z_{nk}=1)\sum_j p(\mathbf{x_n}|z_{nj}=1)} \\ &= \frac{\pi_k^t \mathcal{N}(\mathbf{x}|\mu_k^t,\Sigma_k^t)}{\sum_{j=1}^K \pi_j^t \mathcal{N}(\mathbf{x}|\mu_j^t,\Sigma_j^t)} \end{aligned} \tag{A.16}$$

The M-step then updates the Gaussian parameters given the computed responsibilities:

$$\mu_k^{t+1} = \frac{\sum_n \gamma(z_{nk})\mathbf{x_n}}{\sum_n \gamma(z_{nk})} \tag{A.17}$$

$$\Sigma_k^{t+1} = \frac{\sum_n \gamma(z_{nk})(\mathbf{x_n}-\mu_k^{t+1})(\mathbf{x_n}-\mu_k^{t+1})^T}{\sum_n \gamma(z_{nk})} \tag{A.18}$$

$$\pi_k^{t+1} = \frac{1}{N}\sum_n \gamma(z_{nk}) \tag{A.19}$$

The term $\sum_n \gamma(z_{nk})$ can be interpreted as the number of samples that have been assigned to the distribution k.

In this chapter, the approximate resection cavity area (delineated by the spatial prior $\mathcal{S}_{I,s}$) is modeled very simply as a mixture of 4 Gaussians: two classes for residual tumor and brain, and two classes for the cavity. The cavity is then isolated by selecting the Gaussian with the brightest and darkest means. This results in a probability likelihood for the resection cavity defined as:

$$P(I(\mathbf{x}),J(\mathbf{x})|l_p,s^{l_p}=1) = \frac{1}{2}\mathcal{N}(J(\mathbf{x}+d^{l_p})|\mu_{min},\Sigma_{min}) + \frac{1}{2}\mathcal{N}(J(\mathbf{x}+d^{l_p})|\mu_{max},\Sigma_{max}) \tag{A.20}$$

GMM modeling is illustrated in Fig. [A.1] and the corresponding probability score is shown in Fig. [A.2]

A.4.2 Background area

In the background area, the likelihood term plays the part of the similarity measure that will map the preoperative image to the intraoperative image. The likelihood registration term should take the form of a probabilistic measure that is directly comparable with the resection detection probability. Because of the strong differences in the acquisition process, images of the same modality still have very different intensity profiles. Intensity based similarity measures cannot be used in this situation. A natural probabilistic measure of the differences between the source and target image is to compute the joint intensity histogram between both images. The intensity values of the source and target image are partitioned in a number N of predefined bins $B_I = \{B_I(i)\}_{i\in 1,..,N}, B_J = \{B_J(i)\}_{i\in 1,..,N}$, each bin corresponding to an intensity range. The joint histogram $H(B_I(i), B_J(j))$ is computed by counting the number of voxels in I and J that have a value corresponding to the range defined by bins $B_I(i)$ and $B_J(j)$. When normalized, the joint histogram yields a probability distribution for pairs of intensities:

$$p(I(\mathbf{x}), J(\mathbf{x})) = \frac{H(B_I(i_x), B_J(j_x))}{N_{IJ}} \tag{A.21}$$

where N_{IJ} is the number of intensity pairs counted in the histogram, i_x and j_x are the bin indices associated to intensity values $I(\mathbf{x})$ and $J(\mathbf{x})$ respectively. A sharp histogram is characteristic of well registered images, as the matching intensities for a tissue remain concentrated in a small histogram location. Mutual information and joint entropy measures aim at directly minimizing the histogram dispersion. Maximizing the joint histogram yields the same effect. The source image is deformed so that its intensity pairs with the target image correspond to bin values that have a high probability. In other words, it reduces the histogram's dispersion.

A.5 Experiments

A.5.1 Implementation Details

We adopted a coarse to fine incremental hierarchical approach. We use 4 grid levels and 3 images levels, with 3 iterations at each level. The grid resolution ranged from $8 \times 8 \times 6$ to $57 \times 57 \times 37$. At each iteration, the joint histogram is re-estimated, exploiting the fact that the images should have stronger correspondences as they get registered. In order to avoid introducing a bias due to the high number of voxels in the background, we do not take the background into account when evaluating the joint histogram. The method was tested using the implementation proposed in chapter 4 that does not include computation of registration and segmentation uncertainties. We therefore use a sparse sampling for all iterations and resolution levels, as well as a manually set penalty term defined as $V^t_{pen} = \int_\Omega \bar{\eta}(\|\mathbf{x} - p\|)0.2\mathcal{R}(\mathbf{x})s^{l_p}e^{-l}$ where t is the iteration and l the resolution level. This penalty term had to be manually edited to the different context, which highlights the interest of the uncertainty based propagation.

The number of bins was set at $N = 64$ at the coarsest resolution level and doubled at each subsequent level of the image resolution, so that we have $N = 256$ at the last two grid levels

(maximal image resolution). The last resolution level was dedicated to segmentation as the label set was limited to the segmentation labels. Such a high resolution was not necessary for increased quality of the registration. The smoothing prior weights λ_s and λ_r were set to 1 and 2 respectively. The projection function was defined as cubic B-splines.

Using those parameters, the algorithm runs for about 4 minutes.

A.5.2 Results

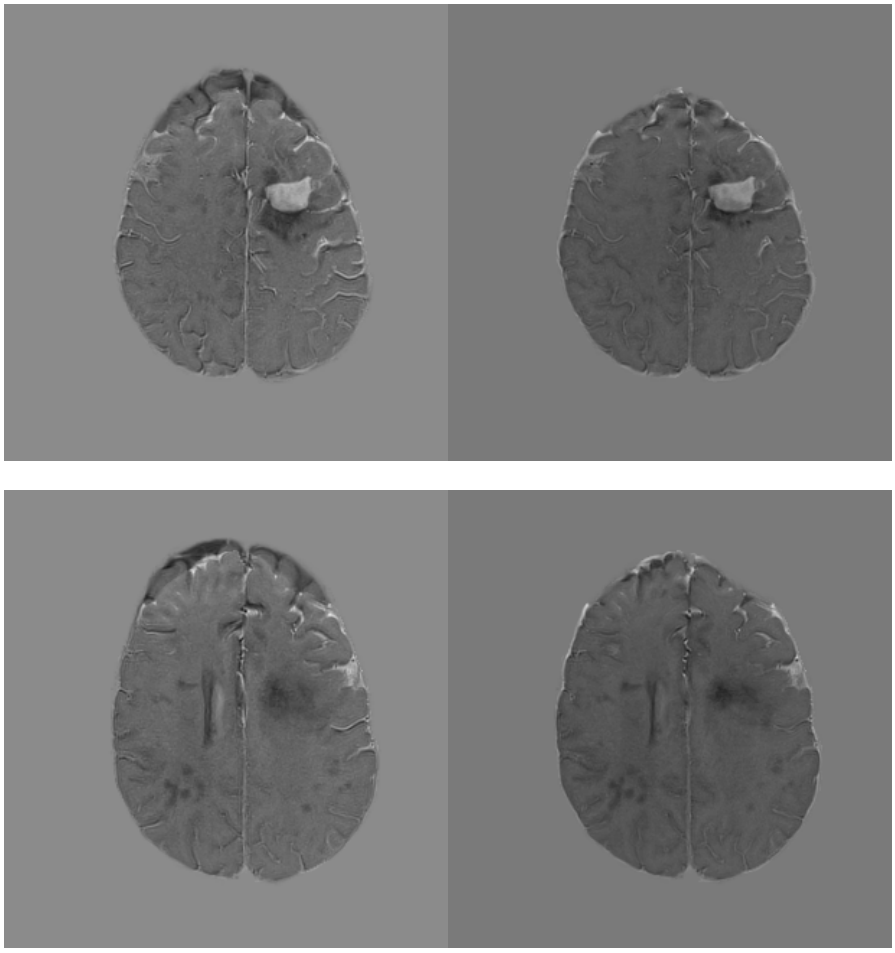
We present preliminary results on a pair of preoperative/intraoperative T2 images. The preoperative tumor, the resection cavity and residual tumor were manually segmented. Both volumes were skullstripped. In order to ensure high correspondences and overlap during the initial estimation of the joint histogram, the preoperative image was affinely registered to the intraoperative image prior to the non-rigid registration using block matching. The preoperative tumor, residual tumor and resection cavity were manually segmented. We compare the automatic resection cavity segmentation obtained with the manual segmentation, and the deformed preoperative tumor to the residual tumor plus resection cavity. Visually, we observe high correspondences between the segmented regions as shown in Fig. [A.4]. Quantitatively, we obtain a dice score of 85 % for residual tumor and 80 % for resection cavity. Strong visual correspondences are observed between the registered preoperative image and the intraoperative image as shown in Fig. [A.3].

A.6 Discussion

This chapter illustrates the adaptability and modularity of the Joint registration and segmentation method introduced in the previous chapter. A probabilistic interpretation of the method is proposed towards registration of preoperative and intraoperative images with tumor resection. The resection cavity is detected through modeling the intensity profile in the resection area as a mixture of Gaussians while a probabilistic similarity criterion for registration is designed based on the images' joint intensity histogram. The method registers the preoperative image to the intraoperative image while segmenting the resection cavity and preserving the surrounding tumor area. This scheme enables direct comparison of the preoperative surgery plan with the current surgery status and gives a direct estimate of the amount of tumor removed and the residual tumor. The speed of the method makes it easily exploitable in a surgery setting, especially given that many preprocessing steps (skullstripping, affine registration) could be carried out preoperatively or at the beginning of the surgery.

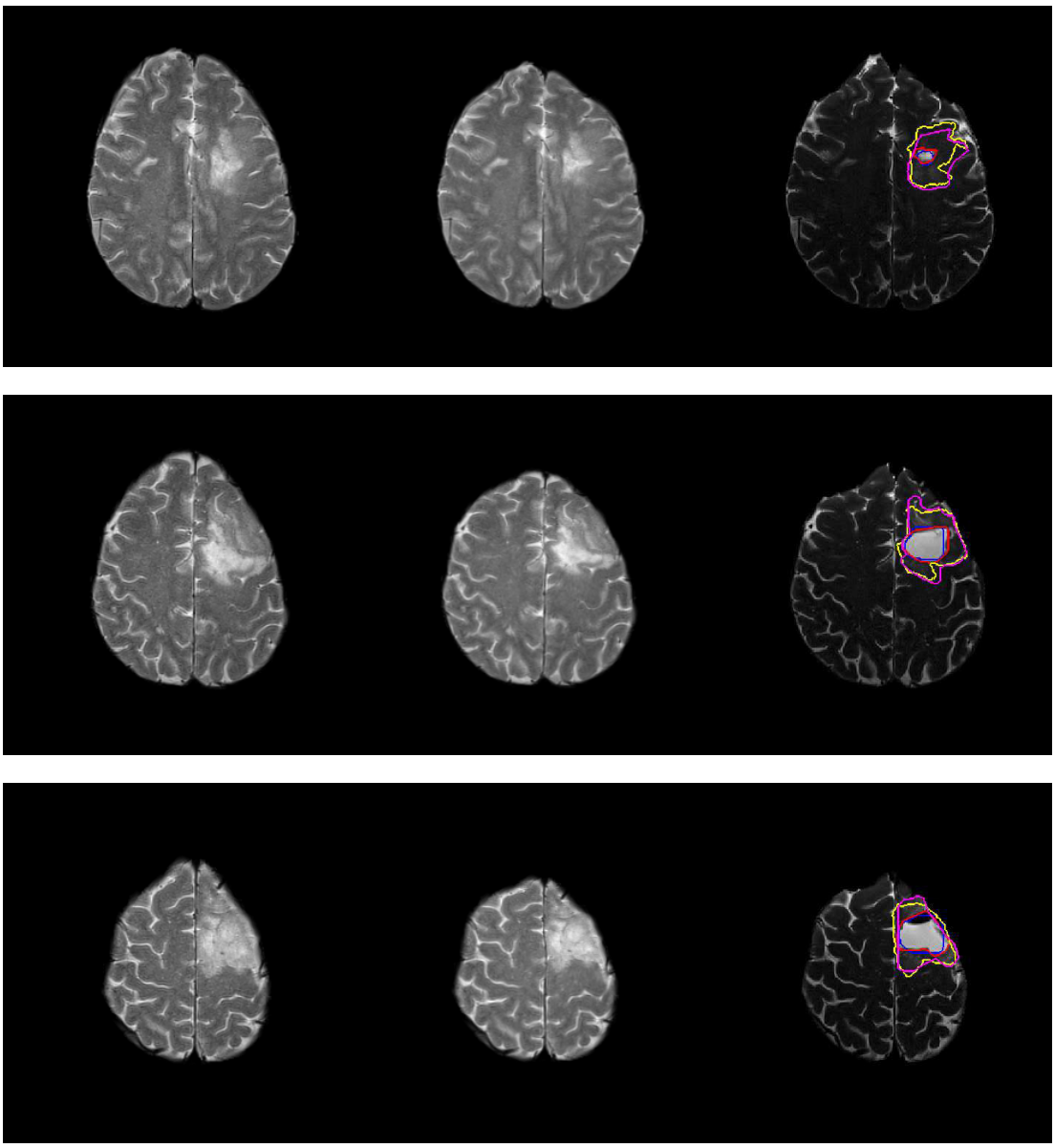
Being tailored for the specific application of tumor resection, the proposed method is less modular in the sense that the matching criterion has to be probabilistic. It is however more adapted to multimodal registration due to the use of the information theoretic based joint histogram as similarity criterion. Given the limited dataset, a training step was not included and the GMM was used to directly model the image's intensity profile and extract Gaussians corresponding to the resection cavity. Learning a GMM directly on a training set of resection cavity voxels would certainly yield

more accurate results, especially in the areas filled with air. Additionally, the method would have to be tested on a larger dataset to thoroughly assess its potential and the possible improvements.



(a) **(b)**

Figure A.3: *Difference Image after affine registration (a) and after non-rigid registration (b).*



(a) (b) (c)

Figure A.4: *Visual registration and segmentation results. (a) Preoperative image after affine registration, (b) non-rigidly deformed preoperative image, (c) intraoperative image segmented. Resection cavity segmentation: automatic (blue) and manual (red); Residual tumor segmentation: manual (magenta) and registered preoperative tumor (yellow).*

Bibliography

[Ackermann 2004] Hermann Ackermann and Axel Riecker. *The contribution of the insula to motor aspects of speech production: a review and a hypothesis*. Brain and language, vol. 89, no. 2, pages 320–328, 2004. 19

[Alchatzidis 2011] Stavros Alchatzidis, Aristeidis Sotiras and Nikos Paragios. *Efficient parallel message computation for MAP inference*. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 1379–1386. IEEE, 2011. 117

[Archip 2007] Neculai Archip, Olivier Clatz, Stephen Whalen, Dan Kacher, Andriy Fedorov, Andriy Kot, Nikos Chrisochoides, Ferenc Jolesz, Alexandra Golby, Peter M Black*et al. Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery*. Neuroimage, vol. 35, no. 2, pages 609–624, 2007. 123

[Ashburner 2005] John Ashburner and Karl J Friston. *Unified segmentation*. Neuroimage, vol. 26, no. 3, pages 839–851, 2005. 91

[Bach Cuadra 2004] Mertixell Bach Cuadra, Claudio Pollo, Anton Bardera, Olivier Cuisenaire, Jean-Guy Villemure and Jean-Philippe Thiran. *Atlas-based segmentation of pathological MR brain images using a model of lesion growth*. Medical Imaging, IEEE Transactions on, vol. 23, no. 10, pages 1301–1314, 2004. 88, 89

[Bach Cuadra 2006] Mertixell Bach Cuadra, Mathieu De Craene, Valérie Duay, Benoî Macq, Claudio Pollo and Jean.Philippe Thiran. *Dense deformation field estimation for atlas-based segmentation of pathological MR brain images*. Computer methods and programs in biomedicine, vol. 84, no. 2-3, pages 66–75, 2006. 89

[Bajcsy 1989] Ruzena Bajcsy and Stane Kovačič. *Multiresolution elastic matching*. Computer vision, graphics, and image processing, vol. 46, no. 1, pages 1–21, 1989. 47

[Bauer 2011] Stefan Bauer, Lutz-P Nolte and Mauricio Reyes. *Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 354–361. Springer, 2011. 42

[Besag 1974] Julian Besag. *Spatial Interaction and the Statistical Analysis of Lattice Systems*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 2, pages 192–236, 1974. 38

[Besag 1986] Julian Besag. *On the statistical analysis of dirty pictures*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 48, pages 259–302, 1986. 39, 40, 92

[Bezdek 1981] James C. Bezdek. Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell, MA, USA, 1981. 31

[Bhatia 2004] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards and D. Rueckert. *Consistent groupwise non-rigid registration for atlas construction*. In Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI, pages 908–911, 2004. 116

[Bishop 2006] Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 1. springer New York, 2006. 128

[Bondiau 2005] Pierre-Yves Bondiau, Grégoire Malandain, Stéphane Chanalet, Pierre-Yves Marcy, Jean-Louis Habrand, Francois Fauchon, Philippe Paquis, Adel Courdi, Olivier Commowick, Isabelle Ruttenet *al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context*. International Journal of Radiation Oncology* Biology* Physics, vol. 61, no. 1, pages 289–298, 2005. 43

[Boser 1992] Bernhard E Boser, Isabelle M Guyon and Vladimir N Vapnik. *A training algorithm for optimal margin classifiers*. In Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152. ACM, 1992. 32

[Boykov 2001] Yuri Boykov, Olga Veksler and Ramin Zabih. *Fast Approximate Energy Minimization via Graph Cuts*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pages 1222–1239, 2001. 40

[Boykov 2004] Yuri Boykov and Vladimir Kolmogorov. *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, no. 9, pages 1124–1137, 2004. 40

[Boykov 2006] Yuri Boykov and Gareth Funka-Lea. *Graph Cuts and Efficient N-D Image Segmentation*. International Journal of Computer Vision, vol. 70, no. 2, pages 109–131, 2006. 37, 70

[Breiman 2001] Leo Breiman. *Random forests*. Machine learning, vol. 45, no. 1, pages 5–32, 2001. 34

[Brett 2001] Matthew Brett, Alexander P Leff, Chris Rorden and John Ashburner. *Spatial normalization of brain images with focal lesions using cost function masking*. Neuroimage, vol. 14, no. 2, pages 486–500, 2001. 87

[Bronstein 2010] Michael M Bronstein, Alexander M Bronstein, Fabrice Michel and Nikos Paragios. *Data fusion through cross-modality metric learning using similarity-sensitive hashing*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 3594–3601. IEEE, 2010. 45

[Brown 1992] Lisa Gottesfeld Brown. *A survey of image registration techniques*. ACM computing surveys (CSUR), vol. 24, no. 4, pages 325–376, 1992. 44

[Buxton 2009] Richard B Buxton. Introduction to functional magnetic resonance imaging: principles and techniques. Cambridge University Press, 2009. 26

[Bynevelt 2001] M. Bynevelt, J. Britton, H. Seymour, E. MacSweeney, N. Thomas and K. Sandhu. *FLAIR imaging in the follow-up of low-grade gliomas: time to dispense with the dual-echo?* Neuroradiology, vol. 43, pages 129–133, Feb 2001. 27

[Cabezas 2011] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet and Meritxell Bach Cuadra. *A review of atlas-based segmentation for magnetic resonance brain images*. Computer methods and programs in biomedicine, vol. 104, no. 3, pages e158–e177, 2011. 50

[Cachier 2003] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec and Nicholas Ayache. *Iconic Feature Based Nonrigid Registration: The PASHA Algorithm*. Comp. Vision and Image Understanding, vol. 89, no. 2-3, pages 272–298, Feb.-march 2003. Special Issue on Nonrigid Registration. 47, 88

[Capelle 2013] Laurent Capelle, Denys Fontaine, Emmanuel Mandonnet, Luc Taillandier, Jean Louis Golmard, Luc Bauchet, Johan Pallud, Philippe Peruzzi, Marie Hélène Baron, Michèle Kujas*et al. Spontaneous and therapeutic prognostic factors in adult hemispheric World Health Organization Grade II gliomas: a series of 1097 cases: Clinical article*. Journal of neurosurgery, vol. 118, no. 6, pages 1157–1168, 2013. 16

[Chan 2001] Tony F Chan and Luminita A Vese. *Active contours without edges*. Image Processing, IEEE Transactions on, vol. 10, no. 2, pages 266–277, 2001. 30

[Chang 2009] Edward F Chang, Aaron Clark, Randy L Jensen, Mark Bernstein, Abhijit Guha, Giorgio Carrabba, Debabrata Mukhopadhyay, Won Kim, Linda M Liau, Susan M Chang*et al. Multiinstitutional validation of the University of California at San Francisco low-grade glioma prognostic scoring system: clinical article*. Journal of neurosurgery, vol. 111, no. 2, pages 203–210, 2009. 18

[Chen 2002] Yunmei Chen, Hemant D Tagare, Sheshadri Thiruvenkadam, Feng Huang, David Wilson, Kaundinya S Gopinath, Richard W Briggs and Edward A Geiser. *Using prior shapes in geometric active contours in a variational framework*. International Journal of Computer Vision, vol. 50, no. 3, pages 315–328, 2002. 28

[Chitphakdithai 2010] Nicha Chitphakdithai and James S Duncan. *Non-rigid registration with missing correspondences in preoperative and postresection brain images*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, pages 367–374. Springer, 2010. 124

[Chitphakdithai 2011] N. Chitphakdithai, K.P. Vives and J.S. Duncan. *Registration of brain resection MRI with intensity and location priors*. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 1520–1523, 2011. 124

[Choi 2000] Yongchoel Choi and Seungyong Lee. *Injectivity conditions of 2D and 3D uniform cubic B-spline functions*. Graphical models, vol. 62, no. 6, pages 411–427, 2000. 47

[Christensen 1994] Gary Edward Christensen. *Deformable shape models for anatomy*. PhD thesis, Washington University Saint Louis, Mississippi, 1994. 47

[Chui 2003] Haili Chui and Anand Rangarajan. *A new point matching algorithm for non-rigid registration*. Computer Vision and Image Understanding, vol. 89, no. 2, pages 114–141, 2003. 123

[Clark 1998] Matthew C. Clark, Lawrence O. Hall, Dmitry B. Goldgof, Robert Velthuizen, F. Reed Murtagh and Martin S. Silbiger. *Automatic tumor segmentation using knowledge-based techniques*. IEEE Transactions on Medical Imaging, vol. 17, no. 2, pages 187–201, 1998. 32

[Clatz 2005] Olivier Clatz, Hervé Delingette, I-F Talos, Alexandra J Golby, Ron Kikinis, Ferenc A Jolesz, Nicholas Ayache and Simon K Warfield. *Robust nonrigid registration to capture brain shift from intraoperative MRI*. Medical Imaging, IEEE Transactions on, vol. 24, no. 11, pages 1417–1427, 2005. 122

[Cobzas 2007] Dana Cobzas, Neil Birkbeck, Mark Schmidt, Martin Jagersand and Albert Murtha. *3D variational brain tumor segmentation using a high dimensional feature set*. In IEEE 11th International Conference on Computer Vision - ICCV, pages 1–8. IEEE, 2007. 51

[Collignon 1995] André Collignon, Frederik Maes, Dominique Delaere, Dirk Vandermeulen, Paul Suetens and Guy Marchal. *Automated multi-modality image registration based on information theory*. In Information processing in medical imaging, volume 3, pages 264–274, 1995. 45

[Collins 1994] D Louis Collins, Peter Neelin, Terrence M Peters and Alan C Evans. *Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space*. Journal of computer assisted tomography, vol. 18, no. 2, pages 192–205, 1994. 49

[Commowick 2005] Olivier Commowick, Radu Stefanescu, Pierre Fillard, Vincent Arsigny, Nicholas Ayache, Xavier Pennec and Grégoire Malandain. *Incorporating statistical mea-*

sures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy. In Medical Image Computing and Computer-Assisted Intervention - MICCAI, pages 927–934. Springer, 2005. 87

[Corso 2008] Jason J Corso, Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha and Alan Yuille. *Efficient multilevel brain tumor segmentation with integrated bayesian model classification*. IEEE Transactions on Medical Imaging, vol. 27, no. 5, pages 629–640, 2008. 36, 37

[Davies 1979] David L. Davies and Donald W. Bouldin. *A Cluster Separation Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, pages 224–227, 1979. 61

[Dawant 1999] Benoit M Dawant, SL Hartmann and S Gadamsetty. *Brain atlas deformation in the presence of large space-occupying tumors*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 589–596. Springer, 1999. 88

[Dempster 1977] Arthur P Dempster, Nan M Laird and Donald B Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, pages 1–38, 1977. 41, 127

[Dengler 1988] Joachim Dengler and Markus Schmidt. *The Dynamic Pyramid a Model for the Motion Analysis with Controlled Continuity*. In Image Analysis and Processing II, pages 337–344. Springer, 1988. 51

[Dennis 1977] John E Dennis Jr and Jorge J Moré. *Quasi-Newton methods, motivation and theory*. SIAM review, vol. 19, no. 1, pages 46–89, 1977. 48

[Desmurget 2007] M. Desmurget, F. Bonnetblanc and H. Duffau. *Contrasting acute and slow-growing lesions: a new door to brain plasticity*. Brain, vol. 130, pages 898–914, Apr 2007. 16, 55

[Dolecek 2012] Therese A Dolecek, Jennifer M Propp, Nancy E Stroup and Carol Kruchko. *CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009*. Neuro-oncology, vol. 14, no. suppl 5, pages v1–v49, 2012. 16

[Droske 2001] Marc Droske, Bernhard Meyer, Martin Rumpf and Carlo Schaller. *An adaptive level set method for medical image segmentation*. In Information Processing in Medical Imaging, pages 416–422. Springer, 2001. 29, 30

[Duan 2005] Kai-Bo Duan and S Sathiya Keerthi. *Which is the best multiclass SVM method? An empirical study*. In Multiple Classifier Systems, pages 278–285. Springer, 2005. 34

[Duda 2001] Richard O. Duda, Peter E. Hart and David G. Stork. Pattern classification. Wiley, 2001. 53

[Duffau 2003] H. Duffau, L. Capelle, D. Denvil, N. Sichez, P. Gatignol, M. Lopes, M. C. Mitchell, J. P. Sichez and R. Van Effenterre. *Functional recovery after surgical resection of low grade gliomas in eloquent brain: hypothesis of brain compensation*. J. Neurol. Neurosurg. Psychiatr., vol. 74, pages 901–907, Jul 2003. 16

[Duffau 2004] H. Duffau and L. Capelle. *Preferential brain locations of low-grade gliomas*. Cancer, vol. 100, pages 2622–2626, Jun 2004. 19, 56

[Duffau 2005] H. Duffau. *Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity*. Lancet Neurol, vol. 4, pages 476–486, Aug 2005. 55

[Duffau 2006] H. Duffau, L. Taillandier, P. Gatignol and L. Capelle. *The insular lobe and brain plasticity: Lessons from tumor surgery*. Clin Neurol Neurosurg, vol. 108, pages 543–548, Sep 2006. 16, 17, 18

[Dunn 1973] J. C. Dunn. *A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters*. Journal of Cybernetics, vol. 3, no. 3, pages 32–57, 1973. 31

[Dunn 1974] J. C. Dunn. *Well separated clusters and optimal fuzzy-partitions*. Journal of Cybernetics, vol. 4, pages 95–104, 1974. 61, 62

[Evans 1993] Alan C Evans, D Louis Collins, SR Mills, ED Brown, RL Kelly and Terry M Peters. *3D statistical neuroanatomical models from 305 MRI volumes*. In Nuclear Science Symposium and Medical Imaging Conference,, pages 1813–1817. IEEE, 1993. 49

[Fletcher-Heath 2001] Lynn M Fletcher-Heath, Lawrence O Hall, Dmitry B Goldgof, F Reed Murtagh*et al.* *Automatic segmentation of non-enhancing brain tumors in magnetic resonance images*. Artificial Intelligence in Medicine, vol. 21, no. 1-3, pages 43–63, 2001. 32

[Fonov 2009] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almli and D. L. Collins. *Unbiased nonlinear average age-appropriate brain templates from birth to adulthood*. NeuroImage, vol. 47, page S102, July 2009. 50

[Fonov 2011] Vladimir Fonov, Alan C Evans, Kelly Botteron, C Robert Almli, Robert C McKinstry and D Louis Collins. *Unbiased average age-appropriate atlases for pediatric studies*. NeuroImage, vol. 54, no. 1, pages 313–327, 2011. 50

[Ford 1962] L. R. Ford and D. R. Fulkerson. Flows in Networks. Princeton University Press, 1962. 40

[Freund 1997] Yoav Freund and Robert E Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. Journal of computer and system sciences, vol. 55, no. 1, pages 119–139, 1997. 34, 64, 66

[Friedman 2000] J. Friedman, T. Hastie and R. Tibshirani. *Additive Logistic Regression: a Statistical View of Boosting*. The Annals of Statistics, vol. 38, no. 2, page 337?407, 2000. 34, 66

[Friston 1995] Karl Friston, John Ashburner, Christopher D Frith, J-B Poline, John D Heather, Richard SJ Frackowiak*et al.* *Spatial registration and normalization of images*. Human brain mapping, vol. 3, no. 3, pages 165–189, 1995. 87

[García 2004] Cristina García and José Alí Moreno. *Kernel based method for segmentation and modeling of magnetic resonance images*. In Advances in Artificial Intelligence–IBERAMIA 2004, pages 636–645. Springer, 2004. 32

[Geman 1984] S. Geman and D. Geman. *Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images*. IEEE Transactions on Pattern Recognition, vol. 6, pages 721–741, 1984. 37

[Geremia 2012] Ezequiel Geremia, Bjoern H Menze and Nicholas Ayache. *Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images*. In MICCAI Challenge on Multimodal Brain Tumor Segmentation, 2012. 34

[Gering 2002] David T. Gering, W. Eric L. Grimson and Ron Kikinis. *Recognizing Deviations from Normalcy for Brain Tumor Segmentation*. In Takeyoshi Dohi and Ron Kikinis, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI, volume 2488 of *Lecture Notes in Computer Science*, pages 388–395. Springer, 2002. 41, 51, 64

[Gibbs 1996] P Gibbs, DL Buckley, SJ Blackband and A Horsman. *Tumour volume determination from MR images by morphological segmentation*. Physics in medicine and biology, vol. 41, no. 11, page 2437, 1996. 21

[Glocker 2008a] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab and Nikos Paragios. *Dense image registration through MRFs and efficient linear programming*. Medical Image Analysis, vol. 12, no. 6, pages 731 – 741, 2008. Special issue on information processing in medical imaging 2007. 37, 48, 94, 98, 99

[Glocker 2008b] Ben Glocker, Nikos Paragios, Nikos Komodakis, Georgios Tziritas and Nassir Navab. *Optical flow estimation with uncertainties through dynamic MRFs.* In CVPR, 2008. 73, 95, 99, 101, 105

[Goldberg 1988] Andrew V Goldberg and Robert E Tarjan. *A new approach to the maximum-flow problem*. Journal of the ACM (JACM), vol. 35, no. 4, pages 921–940, 1988. 40

[Gooya 2011] Ali Gooya, George Biros and Christos Davatzikos. *Deformable Registration of Glioma Images Using EM Algorithm and Diffusion Reaction Modeling*. IEEE Trans. Med. Imaging, vol. 30, no. 2, pages 375–390, 2011. 93

[Gooya 2012] Ali Gooya, Kilian M. Pohl, Michel Bilello, Luigi Cirillo, George Biros, Elias R. Melhem and Christos Davatzikos. *GLISTR: Glioma Image Segmentation and Registration*. IEEE Trans. Med. Imaging, vol. 31, no. 10, pages 1941–1954, 2012. 56, 92, 93

[Gozé 2009] C Gozé, V Rigau, L Gibert, T Maudelonde and H Duffau. *Lack of complete 1p19q deletion in a consecutive series of 12 WHO grade II gliomas involving the insula: a marker of worse prognosis?* Journal of neuro-oncology, vol. 91, no. 1, pages 1–5, 2009. 56

[Haralick 1973] Robert M Haralick, Karthikeyan Shanmugam and Its' Hak Dinstein. *Textural features for image classification*. Systems, Man and Cybernetics, IEEE Transactions on, no. 6, pages 610–621, 1973. 35

[Hartkens 2003] Thomas Hartkens, Derek LG Hill, Andy D Castellano-Smith, David J Hawkes, Calvin R Maurer Jr, Alastair J Martin, Walter A Hall, Haiying Liu and Charles L Truwit. *Measurement and analysis of brain deformation during neurosurgery*. Medical Imaging, IEEE Transactions on, vol. 22, no. 1, pages 82–92, 2003. 121

[Hestenes 1952] Magnus R Hestenes and Eduard Stiefel. *Methods of Conjugate Gradients for Solving Linear Systems1*. Journal of Research of the National Bureau of Standards, vol. 49, no. 6, pages 409–436, 1952. 48

[Ho 2002] Sean Ho, Elizabeth Bullitt and Guido Gerig. *Level-set evolution with region competition: automatic 3-D segmentation of brain tumors*. In 16th International Conference on Pattern Recognition, volume 1, pages 532–535. IEEE, 2002. 30, 51

[Hogea 2007] Cosmina Hogea, George Biros, Feby Abraham and Christos Davatzikos. *A robust framework for soft tissue simulations with application to modeling brain tumor mass effect in 3D MR images*. Physics in medicine and biology, vol. 52, no. 23, page 6893, 2007. 90

[Hogea 2008] Cosmina Hogea, Christos Davatzikos and George Biros. *An image-driven parameter estimation problem for a reaction–diffusion glioma growth model with mass effects*. Journal of mathematical biology, vol. 56, no. 6, pages 793–825, 2008. 93

[Holmes 1998] Colin J Holmes, Rick Hoge, Louis Collins, Roger Woods, Arthur W Toga and Alan C Evans. *Enhancement of MR images using registration for signal averaging*. Journal of computer assisted tomography, vol. 22, no. 2, pages 324–333, 1998. 49

[Holodny 2000] Andrei I Holodny, Michael Schulder, Wen-Ching Liu, Jonathan Wolko, Joseph A Maldjian and Andrew J Kalnin. *The effect of brain tumors on BOLD functional MR imaging activation in the adjacent motor cortex: implications for image-guided neurosurgery*. American journal of neuroradiology, vol. 21, no. 8, pages 1415–1422, 2000. 17

[Ius 2011] Tamara Ius, Elsa Angelini, Michel Thiebaut de Schotten, Emmanuel Mandonnet and Hugues Duffau. *Evidence for potentials and limitations of brain plasticity using an atlas of functional resectability of WHO grade II gliomas: towards a ¡Èminimal common brain¡É*. Neuroimage, vol. 56, no. 3, pages 992–1000, 2011. 55

[Jakola 2012] Asgeir S Jakola, Kristin S Myrmel, Roar Kloster, Sverre H Torp, Sigurd Lindal, Geirmund Unsgård and Ole Solheim. *Comparison of a Strategy Favoring Early Surgical Resection vs a Strategy Favoring Watchful Waiting in Low-Grade GliomasSurgical Resection vs Waiting in Low-Grade Gliomas*. JAMA, vol. 308, no. 18, pages 1881–1888, 2012. 16

[Jbabdi 2005] S. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K. R. Swanson, M. Pelegrini-Issac, R. Guillevin and H. Benali. *Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging*. Magn Reson Med, vol. 54, pages 616–624, Sep 2005. 116

[Johnson 1967] Stephen C Johnson. *Hierarchical clustering schemes*. Psychometrika, vol. 32, no. 3, pages 241–254, 1967. 59

[Joshi 2004] S. Joshi, B. Davis, M. Jomier and G. Gerig. *Unbiased diffeomorphic atlas construction for computational anatomy*. NeuroImage, vol. 23, pages S151–S160, 2004. 116

[Karantzalos 2005] Konstantinos Karantzalos and Nikos Paragios. *Implicit free-form-deformations for multi-frame segmentation and tracking*. In Variational, Geometric, and Level Set Methods in Computer Vision, pages 271–282. Springer, 2005. 91

[Kass 1988] Michael Kass, Andrew Witkin and Demetri Terzopoulos. *Snakes: Active contour models*. International Journal of Computer Vision, vol. 1, no. 4, pages 321–331, 1988. 28

[Kaus 1999] MR Kaus, Simon K Warfield, Arya Nabavi, E Chatzidakis, Peter M. Black, Ferenc A Jolesz and Ron Kikinis. *Segmentation of meningiomas and low grade gliomas in MRI*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, volume 2, pages 1–10. Springer, 09 1999. 50

[Kaus 2001] M.R. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz and R. Kikinis. *Automated Segmentation of MR Images of Brain Tumors*. Radiology, vol. 218, no. 2, pages 586–591, 02 2001. 18, 50, 85, 88

[Kikinis 1996] Ron Kikinis, Martha Elizabeth Shenton, Dan V. Iosifescu, Robert W. McCarley, Pairash Saiviroonporn, Hiroto H. Hokama, Andre Robatino, David Metcalf, Cynthia G Wible, Chiara M. Portas*et al.* *A digital brain atlas for surgical planning, model-driven segmentation, and teaching*. Visualization and Computer Graphics, IEEE Transactions on, vol. 2, no. 3, pages 232–241, 1996. 49

[Kim 2004] Jeongtae Kim and Jeffrey A Fessler. *Intensity-based image registration using robust correlation coefficients*. IEEE Transactions on Medical Imaging, vol. 23, no. 11, pages 1430–1444, 2004. 44

[Kirkpatrick 1983] Scott Kirkpatrick and Mario P Vecchi. *Optimization by simmulated annealing*. science, vol. 220, no. 4598, pages 671–680, 1983. 40

[Klein 2007] Stefan Klein, Marius Staring and Josien PW Pluim. *Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines*. Image Processing, IEEE Transactions on, vol. 16, no. 12, pages 2879–2890, 2007. 48

[Kohli 2007] Pushmeet Kohli and Philip HS Torr. *Dynamic graph cuts for efficient inference in markov random fields*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 12, pages 2079–2088, 2007. 100

[Kohli 2008] Pushmeet Kohli and Philip H. S. Torr. *Measuring uncertainty in graph cut solutions*. Computer Vision and Image Understanding, vol. 112, no. 1, pages 30–38, October 2008. 99, 101

[Kolmogorov 2002] Vladimir Kolmogorov and Ramin Zabih. *Multi-camera Scene Reconstruction via Graph Cuts*. In European Conference on Computer Vision-ECCV, pages 82–96, London, UK, UK, 2002. Springer-Verlag. 37

[Kolmogorov 2006] Vladimir Kolmogorov. *Convergent tree-reweighted message passing for energy minimization*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 10, pages 1568–1583, 2006. 41

[Kolmogorov 2007] Vladimir Kolmogorov and Carsten Rother. *Minimizing nonsubmodular functions with graph cuts-a review*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 7, pages 1274–1279, 2007. 40

[Komodakis 2007a] Nikos Komodakis, Nikos Paragios and Georgios Tziritas. *MRF optimization via dual decomposition: Message-passing revisited*. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007. 41

[Komodakis 2007b] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Fast, Approximately Optimal Solutions for Single and Dynamic MRFs*. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 0, pages 1–8, 2007. 37, 40, 70

[Komodakis 2008a] Nikos Komodakis, Nikos Paragios and Georgios Tziritas. *Clustering via LP-based Stabilities*. In NIPS, pages 865–872, 2008. 60

[Komodakis 2008b] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies*. Comput. Vis. Image Underst., vol. 112, no. 1, pages 14–29, 2008. 40, 70, 71, 96

[Konukoglu 2010] Ender Konukoglu, Olivier Clatz, Bjoern H Menze, Bram Stieltjes, M-A Weber, Emmanuel Mandonnet, Hervé Delingette and Nicholas Ayache. *Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations*. IEEE Transactions on Medical Imaging, vol. 29, no. 1, pages 77–95, 2010. 116

[Kushner 2003] Harold J Kushner and George Yin. Stochastic approximation and recursive algorithms and applications, volume 35. Springer, 2003. 48

[Kyriacou 1999] Stelios K. Kyriacou, Christos Davatzikos, S. James Zinreich and R Nick Bryan. *Nonlinear elastic registration of brain images with tumor pathology using a biomechanical model [MRI]*. IEEE Transactions on Medical Imaging, vol. 18, no. 7, pages 580–592, 1999. 89

[Lafferty 2001] John D. Lafferty, Andrew McCallum and Fernando C. N. Pereira. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. In Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. 39

[Lamecker 2010] Hans Lamecker and Xavier Pennec. *Atlas to image-with-tumor registration based on demons and deformation inpainting*. In Proc. MICCAI Workshop on Computational Imaging Biomarkers for Tumors-From Qualitative to Quantitative, CIBT. Citeseer. Citeseer, 2010. 87, 88

[Larjavaara 2007] S. Larjavaara, R. Mantyla, T. Salminen, H. Haapasalo, J. Raitanen, J. Jaaskelainen and A. Auvinen. *Incidence of gliomas by anatomic location*. Neuro-oncology, vol. 9, pages 319–325, Jul 2007. 56

[Lee 2005] Chi-Hoon Lee, Mark Schmidt, Albert Murtha, Aalo Bistritz, Jöerg Sander and Russell Greiner. *Segmenting brain tumors with conditional random fields and support vector machines*. In Computer Vision for Biomedical Image Applications, pages 469–478. Springer, 2005. 42

[Lee 2008] Chi-Hoon Lee, Shaojun Wang, Albert Murtha, Matthew RG Brown and Russell Greiner. *Segmenting brain tumors using pseudo–conditional random fields*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 359–366. Springer, 2008. 42

[Lee 2009] Daewon Lee, Matthias Hofmann, Florian Steinke, Yasemin Altun, Nathan D Cahill and Bernhard Scholkopf. *Learning similarity measure for multi-modal 3D image registration*. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pages 186–193. IEEE, 2009. 45

[Lefohn 2003] Aaron E Lefohn, Joshua E Cates and Ross T Whitaker. *Interactive, GPU-based level sets for 3D segmentation*. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003, pages 564–572. Springer, 2003. 29

[Leventon 2000] Michael E Leventon, W Eric L Grimson and Olivier Faugeras. *Statistical shape influence in geodesic active contours*. In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 1, pages 316–323. IEEE, 2000. 28, 29

[Liu 2005] Jianguo Liu, Jayaram K Udupa, Dewey Odhner, David Hackney and Gul Moonis. *A system for brain tumor volume estimation via MR imaging and fuzzy connectedness.* Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society, vol. 29, no. 1, page 21, 2005. 36

[Liu 2010] Yixun Liu, Chengjun Yao, LiangFu Zhou and Nikos Chrisochoides. *A point based non-rigid registration for tumor resection using IMRI*. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 1217–1220. IEEE, 2010. 123

[Louis 2007] David N Louis, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvet, Bernd W Scheithauer and Paul Kleihues. *The 2007 WHO classification of tumours of the central nervous system*. Acta neuropathologica, vol. 114, no. 2, pages 97–109, 2007. 15

[MacQueen 1967] James MacQueen*et al.* *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, page 14. California, USA, 1967. 30, 59

[Maes 1997] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal and Paul Suetens. *Multimodality image registration by maximization of mutual information*. Medical Imaging, IEEE Transactions on, vol. 16, no. 2, pages 187–198, 1997. 45

[Mahapatra 2012] Dwarikanath Mahapatra and Ying Sun. *Integrating segmentation information for improved MRF-based elastic image registration*. IEEE Transactions on Medical Imaging, vol. 21, no. 1, pages 170–183, 2012. 92

[Maintz 2001] JB Antoine Maintz, Petra A van den Elsen and Max A Viergever. *3D multimodality medical image registration using morphological tools*. Image and vision computing, vol. 19, no. 1, pages 53–62, 2001. 45

[Malladi 1995] Ravi Malladi, James A. Sethian and Baba C. Vemuri. *Shape modeling with front propagation: A level set approach*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 2, pages 158–175, 1995. 28

[Mandonnet 2003] E. Mandonnet, J. Y. Delattre, M. L. Tanguy, K. R. Swanson, A. F. Carpentier, H. Duffau, P. Cornu, R. Van Effenterre, E. C. Alvord and L. Capelle. *Continuous growth of mean tumor diameter in a subset of grade II gliomas*. Ann. Neurol., vol. 53, pages 524–528, Apr 2003. 15, 17, 19

[Mandonnet 2008] E. Mandonnet, J. Pallud, O. Clatz, L. Taillandier, E. Konukoglu, H. Duffau and L. Capelle. *Computational modeling of the WHO grade II glioma dynamics: principles and applications to management paradigm*. Neurosurg Rev, vol. 31, pages 263–269, Jul 2008. 17, 19

[Manjunath 1996] Bangalore S Manjunath and Wei-Ying Ma. *Texture features for browsing and retrieval of image data*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 18, no. 8, pages 837–842, 1996. 35, 67

[Mazziotta 2001] John Mazziotta, Arthur Toga, Alan Evans, Peter Fox, Jack Lancaster, Karl Zilles, Roger Woods, Tomas Paus, Gregory Simpson, Bruce Pike*et al.* *A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)*. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, vol. 356, no. 1412, pages 1293–1322, 2001. 50

[Menze 2010] Bjoern H Menze, Koen Van Leemput, Danial Lashkari, Marc-André Weber, Nicholas Ayache and Polina Golland. *A generative model for brain tumor segmentation in multi-modal images*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 151–159. Springer, 2010. 53, 64

[Michel 2011] F. Michel, M. Bronstein, A. Bronstein and N. Paragios. *Boosted metric learning for 3D multi-modal deformable registration*. In ISBI, pages 1209–1214, 2011. 67

[Mohamed 2005] Ashraf Mohamed and Christos Davatzikos. *Finite element modeling of brain tumor mass-effect from 3D medical images*. In Medical Image Computing and Computer-Assisted Intervention - MICCAI, pages 400–408. Springer, 2005. 89

[Mohamed 2006] Ashraf Mohamed, Evangelia I Zacharaki, Dinggang Shen and Christos Davatzikos. *Deformable registration of brain tumor images via a statistical model of tumor-induced deformation*. Medical image analysis, vol. 10, no. 5, pages 752–763, 2006. 89

[Moon 2002] Nathan Moon, Elizabeth Bullitt, Koen Van Leemput and Guido Gerig. *Model-Based Brain and Tumor Segmentation*. In ICPR (1), pages 528–531, 2002. 51

[Moonis 2002] Gul Moonis, Jianguo Liu, Jayaram K Udupa and David B Hackney. *Estimation of tumor volume with fuzzy-connectedness segmentation of MR images*. American Journal of Neuroradiology, vol. 23, no. 3, pages 356–363, 2002. 36

[Nyúl 1999] László G Nyúl and Jayaram K Udupa. *On standardizing the MR image intensity scale*. Magnetic resonance in medicine, vol. 42, pages 1072–1081, 1999. 27

[Ogawa 1990] Seiji Ogawa, TM Lee, AR Kay and DW Tank. *Brain magnetic resonance imaging with contrast dependent on blood oxygenation*. Proceedings of the National Academy of Sciences, vol. 87, no. 24, pages 9868–9872, 1990. 23

[Osher 1988] Stanley Osher and James A Sethian. *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*. Journal of computational physics, vol. 79, no. 1, pages 12–49, 1988. 28

[Ourselin 2000] Sébastien Ourselin, Alexis Roche, Sylvain Prima and Nicolas Ayache. *Block matching: a general framework to improve robustness of rigid registration of medical images*. In S.L. Delp, A.M. DiGioia and B. Jaramaz, editeurs, 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI, volume 1935 of *Lecture Notes in Computer Science*, pages 557–566, Pittsburgh, USA, October 2000. Springer. 57, 58, 123

[Pal 1997] Nikhil R Pal and J Biswas. *Cluster validation using graph theoretic concepts*. Pattern Recognition, vol. 30, no. 6, pages 847–857, 1997. 62

[Pallud 2012] Johan Pallud, Luc Taillandier, Laurent Capelle, Denys Fontaine, Matthieu Peyre, François Ducray, Hugues Duffau and Emmanuel Mandonnet. *Quantitative morphological magnetic resonance imaging follow-up of low-grade glioma: a plea for systematic measurement of growth rates*. Neurosurgery, vol. 71, no. 3, pages 729–740, 2012. 18

[Paragios 2002] Nikos Paragios and Rachid Deriche. *Geodesic active regions and level set methods for supervised texture segmentation*. International Journal of Computer Vision, vol. 46, no. 3, pages 223–247, 2002. 30

[Paragios 2005] Nikos Paragios and Rachid Deriche. *Geodesic active regions and level set methods for motion estimation and tracking*. Computer Vision and Image Understanding, vol. 97, no. 3, pages 259–282, 2005. 91

[Parisot] Sarah Parisot, William M. Wells III, Stéphane Chemouny, Hugues Duffau and Nikos Paragios. Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. Medical Image Analysis (in press).

[Parisot 2011] Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Graph Based Spatial Position Mapping of Low-Grade Gliomas*. In G. Fichtinger, A.Martel, and T. Peters: MICCAI, Part II, LNCS 6892, pp. 508-515. Springer, Heidelberg, 2011. 56

[Parisot 2012a] Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Graph-based detection, segmentation & characterization of brain tumors*. In IEEE Conference on Computer Vision and Pattern Recognition - CVPR, pages 988–995. IEEE, 2012.

[Parisot 2012b] Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Joint tumor segmentation and dense deformable registration of brain MR images*. In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editeurs, MICCAI (2), volume 7511, pages 651–658. Springer, 2012. 98

[Parisot 2013] Sarah Parisot, William Wells III, Stephane Chemouny, Hugues Duffau and Nikos Paragios. *Uncertainty-Driven Efficiently-Sampled Sparse Graphical Models for Concurrent Tumor Segmentation and Atlas Registration*. In The IEEE International Conference on Computer Vision (ICCV), December 2013.

[Pearl 1988] Judea Pearl. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, 1988. 41

[Pennec 1999] Xavier Pennec, Pascal Cachier and Nicholas Ayache. *Understanding the demon¡Çs algorithm: 3D non-rigid registration by gradient descent*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 597–605. Springer, 1999. 47, 87

[Phillips 1995] WE Phillips, RP Velthuizen, S Phuphanich, LO Hall, LP Clarke and ML Silbiger. *Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme*. Magnetic resonance imaging, vol. 13, no. 2, pages 277–290, 1995. 31

[Pluim 2003] Josien PW Pluim, JB Antoine Maintz and Max A Viergever. *Mutual-information-based registration of medical images: a survey*. Medical Imaging, IEEE Transactions on, vol. 22, no. 8, pages 986–1004, 2003. 45

[Pohl 2006] Kilian M Pohl, John Fisher, W Eric L Grimson, Ron Kikinis and William M Wells. *A Bayesian model for joint segmentation and registration*. NeuroImage, vol. 31, no. 1, pages 228–239, 2006. 92, 124

[Pollo 2005] Claudio Pollo, Meritxell Bach Cuadra, Olivier Cuisenaire, Jean-Guy Villemure and Jean-Philippe Thiran. *Segmentation of brain structures in presence of a space-occupying lesion*. Neuroimage, vol. 24, no. 4, pages 990–996, 2005. 88

[Popuri 2012] Karteek Popuri, Dana Cobzas, Albert Murtha and Martin Jägersand. *3D variational brain tumor segmentation using Dirichlet priors on a clustered feature set*. International journal of computer assisted radiology and surgery, vol. 7, no. 4, pages 493–506, 2012. 30

[Prastawa 2003] Marcel Prastawa, Elizabeth Bullit, Nathan Moon, Koen Van Leemput and Guido Gerig. *Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors*. Academic Radiology, vol. 10, no. 12, pages 1341–1348, 2003. 51, 52, 64

[Prastawa 2004] Marcel Prastawa, Elizabeth Bullitt, Sean Ho and Guido Gerig. *A brain tumor segmentation framework based on outlier detection*. Medical Image Analysis, vol. 8, no. 3, pages 275–283, 2004. 52, 64, 85

[Prima 2002] Sylvain Prima, Sébastien Ourselin and Nicholas Ayache. *Computation of the mid-sagittal plane in 3-D brain images*. Medical Imaging, IEEE Transactions on, vol. 21, no. 2, pages 122–138, 2002. 77

[Puddephat 2002] Mike Puddephat. *Principles of magnetic resonance imaging*. Easy Measure, 2002. 25

[Rekik 2013] Islem Rekik, Stéphanie Allassonnière, Olivier Clatz, Ezequiel Geremia, Erin Stretton, Hervé Delingette and Nicholas Ayache. *Tumor growth parameters estimation and source localization from a unique time point: Application to low-grade gliomas*. Computer Vision and Image Understanding, vol. 117, no. 3, pages 238 – 249, 2013. 116, 117

[Ren 2012] Xiaohui Ren, Xiangli Cui, Song Lin, Junmei Wang, Zhongli Jiang, Dali Sui, Jing Li and Zhongcheng Wang. *Co-deletion of chromosome 1p/19q and IDH1/2 mutation in glioma subsets of brain tumors in Chinese patients*. PloS one, vol. 7, no. 3, page e32764, 2012. 56

[Rigau 2011] Valérie Rigau, Sonia Zouaoui, Hélène Mathieu-Daudé, Amélie Darlix, Aurélie Maran, Brigitte Trétarre, Faiza Bessaoud, Fabienne Bauchet, Redha Attaoua, Pascale Fabbro-Peray*et al.* *French Brain Tumor DataBase: 5-Year Histological Results on 25 756 Cases*. Brain Pathology, vol. 21, no. 6, pages 633–644, 2011. 16

[Risholm 2009] Petter Risholm, Eigil Samset, Ion-Florin Talos and William M. Wells III. *A Non-rigid Registration Framework That Accommodates Resection and Retraction*. In IPMI, pages 447–458, 2009. 123

[Risholm 2010] Petter Risholm, Eigil Samset and William Wells III. *Validation of a nonrigid registration framework that accommodates tissue resection*. In SPIE Medical Imaging, pages 762319–762319. International Society for Optics and Photonics, 2010. 123

[Roche 1998] Alexis Roche, Grégoire Malandain, Xavier Pennec and Nicholas Ayache. *The correlation ratio as a new similarity measure for multimodal image registration*. In Medical Image Computing and Computer-Assisted Interventation¡ $\frac{1}{2}$MICCAI, pages 1115–1124. Springer, 1998. 44

[Roche 2001] Alexis Roche, Xavier Pennec, Grégoire Malandain and Nicholas Ayache. *Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information*. Medical Imaging, IEEE Transactions on, vol. 20, no. 10, pages 1038–1049, 2001. 45

[Rousseeuw 1987] P. J. Rousseeuw. *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics, vol. 20, pages 53–65, 1987. 61

[Rousseeuw 1999] Peter J Rousseeuw and Katrien Van Driessen. *A fast algorithm for the minimum covariance determinant estimator*. Technometrics, vol. 41, no. 3, pages 212–223, 1999. 52

[Rousson 2002] Mikael Rousson and Nikos Paragios. *Shape priors for level set representations*. In ECCV, pages 78–92. Springer, 2002. 28

[Rueckert 1999] D. Rueckert, L. Sonoda, I. Hayes, D. Hill, M. Leach and D. Hawkes. *Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images*. IEEE Transactions on Medical Imaging, vol. 18, no. 8, pages 712–721, 08 1999. 47, 94

[Sanai 2008] Nader Sanai and Mitchel S Berger. *Glioma extent of resection and its impact on patient outcome*. Neurosurgery, vol. 62, no. 4, pages 753–766, 2008. 16

[Sanai 2011] Nader Sanai, Susan Chang and Mitchel S Berger. *Low-grade gliomas in adults: a review*. Journal of neurosurgery, vol. 115, no. 5, pages 948–965, 2011. 21

[Sederberg 1986] Thomas W Sederberg and Scott R Parry. *Free-form deformation of solid geometric models*. In ACM Siggraph Computer Graphics, volume 20, pages 151–160. ACM, 1986. 47, 94

[Sethian 1996] James A Sethian. *A fast marching level set method for monotonically advancing fronts*. Proceedings of the National Academy of Sciences, vol. 93, no. 4, pages 1591–1595, 1996. 29

[Sharon 2006] Eitan Sharon, Meirav Galun, Dahlia Sharon, Ronen Basri and Achi Brandt. *Hierarchy and adaptivity in segmenting visual scenes*. Nature, vol. 442, no. 7104, pages 810–813, 2006. 37

[Shekhovtsov 2008] Alexander Shekhovtsov, Ivan Kovtun and Václav Hlaváč. *Efficient MRF deformation model for non-rigid image matching*. Computer Vision and Image Understanding, vol. 112, no. 1, pages 91–99, 2008. 48

[Shen 2002] Dinggang Shen and Christos Davatzikos. *HAMMER: hierarchical attribute matching mechanism for elastic registration*. Medical Imaging, IEEE Transactions on, vol. 21, no. 11, pages 1421–1439, 2002. 44, 89

[Shi 2000] Jianbo Shi and Jitendra Malik. *Normalized Cuts and Image Segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pages 888–905, August 2000. 36

[Shi 2012] Wenzhe Shi, Xiahai Zhuang, Luis Pizarro, Wenjia Bai, Haiyan Wang, Kai-Pin Tung, Philip J. Edwards and Daniel Rueckert. *Registration Using Sparse Free-Form Deformations.* In Nicholas Ayache, Hervé Delingette, Polina Golland and Kensaku Mori, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI, volume 7511, pages 659–666. Springer, 2012. 98

[Sled 1998] John G Sled, Alex P Zijdenbos and Alan C Evans. *A nonparametric method for automatic correction of intensity nonuniformity in MRI data*. Medical Imaging, IEEE Transactions on, vol. 17, no. 1, pages 87–97, 1998. 27

[Soffietti 2010] R Soffietti, B G Baumert, L Bello, A von Deimling, H Duffau, M Frénay, W Grisold, R Grant, F Graus, K Hoang-Xuan, M Klein, B Melin, J Rees, T Siegal, A Smits, R Stupp and W and Wick. *Guidelines on management of low-grade gliomas: report of an EFNS-EANO* Task Force.* Eur J Neurol, vol. 17, no. 9, pages 1124–33, 2010. 16, 23

[Sotiras 2009] Aristeidis Sotiras, Nikos Komodakis, Ben Glocker, Jean Francois Deux and Nikos Paragios. *Graphical Models and Deformable Diffeomorphic Population Registration Using Global and Local Metrics*. In MICCAI (1), pages 672–679, 2009. 116

[Sotiras 2013] A. Sotiras, C. Davatzikos and N. Paragios. *Deformable Medical Image Registration: A Survey*. IEEE Transactions on Medical Imaging, vol. 32, no. 7, pages 1153–1190, 2013. 43

[Stefanescu 2004] Radu Stefanescu, Olivier Commowick, Grégoire Malandain, Pierre-Yves Bondiau, Nicholas Ayache and Xavier Pennec. *Non-rigid atlas to subject registration with pathologies for conformal brain radiotherapy*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 704–711. Springer, 2004. 87

[Taheri 2010] S Taheri, SH Ong and VFH Chong. *Level-set segmentation of brain tumors using a threshold-based speed function*. Image and Vision Computing, vol. 28, no. 1, pages 26–37, 2010. 30

[Talairach 1988] J. Talairach and P. Tournoux. Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging (Thieme Classics). Thieme, January 1988. 49

[Tarlow 2012] Daniel Tarlow and Ryan Prescott Adams. *Revisiting uncertainty in graph cut solutions*. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2440–2447. IEEE, 2012. 99

[Thirion 1998] J. P. Thirion. *Image matching as a diffusion process: an analogy with Maxwell's demons*. Medical Image Analysis, vol. 2, no. 3, pages 243–260, September 1998. 47

[Thompson 2000] Paul Thompson, Michael S Mega and Arthur W Toga. *Disease-specific probabilistic brain atlases*. In Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on, pages 227–234. IEEE, 2000. 55

[Toga 2001] A. W. Toga and P. M. Thompson. *Maps of the brain*. Anat Rec, vol. 265, no. 2, pages 37–53, 2001. 50

[Torresani 2008] Lorenzo Torresani, Vladimir Kolmogorov and Carsten Rother. *Feature correspondence via graph matching: Models and global optimization*. In Computer Vision–ECCV 2008, pages 596–609. Springer, 2008. 76

[Tu 2005] Zhuowen Tu. *Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering*. In Tenth IEEE International Conference on Computer Vision ICCV, volume 2, pages 1589–1596. IEEE, 2005. 42

[Van Leemput 1999] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen and Paul Suetens. *Automated model-based tissue classification of MR images of the brain*. IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 897–908, 1999. 53

[Van Leemput 2001] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester and Paul Suetens. *Automated segmentation of multiple sclerosis lesions by model outlier detection*. IEEE Transactions on Medical Imaging, vol. 20, no. 8, pages 677–688, 2001. 51, 53

[Vercauteren 2008] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Diffeomorphic Demons: Efficient Non-parametric Image Registration.* NeuroImage, vol. 45, no. 1, pages S61–S72, 2008. 47, 87

[Verma 2008] Ragini Verma, Evangelia I Zacharaki, Yangming Ou, Hongmin Cai, Sanjeev Chawla, Seung-Koo Lee, Elias R Melhem, Ronald Wolf and Christos Davatzikos. *Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images*. Academic radiology, vol. 15, no. 8, page 966, 2008. 32

[Vinitski 1997] Simon Vinitski, Carlos Gonzalez, Feroze Mohamed, Tad Iwanaga, Robert L Knobler, Kamil Khalili and John Mack. *Improved intracranial lesion characterization by tissue segmentation based on a 3D feature map*. Magnetic resonance in medicine, vol. 37, no. 3, pages 457–469, 1997. 32

[Viola 1997] Paul Viola and William M Wells III. *Alignment by maximization of mutual information*. International journal of computer vision, vol. 24, no. 2, pages 137–154, 1997. 45

[Vovk 2007] Uros Vovk, Franjo Pernus and Bostjan Likar. *A review of methods for correction of intensity inhomogeneity in MRI*. Medical Imaging, IEEE Transactions on, vol. 26, no. 3, pages 405–421, 2007. 27

[Warfield 2000] Simon K Warfield, Michael Kaus, Ferenc A Jolesz and Ron Kikinis. *Adaptive, template moderated, spatially varying statistical classification*. Medical Image Analysis, vol. 4, no. 1, pages 43–55, 2000. 50, 88

[Wei 2007] Corie W Wei, Gang Guo and David J Mikulis. *Tumor effects on cerebral white matter as characterized by diffusion tensor tractography*. The Canadian Journal of Neurological Sciences, vol. 34, no. 1, pages 62–68, 2007. 23

[Wein 2008] Wolfgang Wein, Shelby Brunke, Ali Khamene, Matthew R Callstrom and Nassir Navab. *Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention*. Medical image analysis, vol. 12, no. 5, pages 577–585, 2008. 45

[Wells III 1996a] William M Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima and Ron Kikinis. *Multi-modal volume registration by maximization of mutual information*. Medical image analysis, vol. 1, no. 1, pages 35–51, 1996. 45

[Wells III 1996b] Williams M Wells III, W Eric L Grimson, Ron Kikinis and Ferenc A Jolesz. *Adaptive segmentation of MRI data*. Medical Imaging, IEEE Transactions on, vol. 15, no. 4, pages 429–442, 1996. 41, 92

[Wels 2008] Michael Wels, Gustavo Carneiro, Alexander Aplas, Martin Huber, Joachim Hornegger and Dorin Comaniciu. *A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-D MRI*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 67–75. Springer, 2008. 42

[Whittle 2010] IR Whittle. *What is the place of conservative management for adult supratentorial low-grade glioma?* In Low-Grade Gliomas, pages 65–79. Springer, 2010. 16

[Wu 1993] Z. Wu and R. Leahy. *An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pages 1101–1113, November 1993. 36

[Wyatt 2003] Paul P Wyatt and J Alison Noble. *MAP MRF joint segmentation and registration of medical images*. Medical Image Analysis, vol. 7, no. 4, pages 539–52, 2003. 91

[Xuan 2007] Xiao Xuan and Qingmin Liao. *Statistical structure analysis in MRI brain tumor segmentation*. In 4th International Conference on Image and Graphics, pages 421–426. IEEE, 2007. 34, 35

[Yezzi 2003] Anthony Yezzi, Lilla Zollei and Tina Kapur. *A variational framework for integrating segmentation and registration through active contours*. In Medical Image Analysis, volume 7, pages 171–185, 2003. 91

[Zacharaki 2008] E.I. Zacharaki, D. Shen, S.K. Lee and C. Davatzikos. *ORBIT: a multiresolution framework for deformable registration of brain tumor images*. IEEE Transations on Medical Imaging, vol. 27, no. 8, pages 1003–1017, 2008. 89, 90

[Zacharaki 2009] Evangelia I. Zacharaki, Cosmina Hogea, Dinggang Shen, George Biros and Christos Davatzikos. *Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth*. NeuroImage, vol. 46, no. 3, pages 762–774, 2009. 90, 92

[Zhan 2003] Yiqiang Zhan and Dinggang Shen. *Automated segmentation of 3D US prostate images using statistical texture-based matching method*. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003, pages 688–696. Springer, 2003. 67

[Zhang 2004] Jianguo Zhang, Kai-Kuang Ma, Meng-Hwa Er, Vincent Chong*et al. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine*. In International Workshop on Advanced Image Technology (IWAIT'04), pages 207–211, 2004. 32

[Zhu 1996] Song Chun Zhu and Alan Yuille. *Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 18, no. 9, pages 884–900, 1996. 30

[Zikic 2012] Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, C Demiralp, J Shotton, OM Thomas, T Das, R Jena and SJ Price. *Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI, pages 369–376. Springer, 2012. 34, 35

[Zollei 2005] L. Zollei, E. Learned Miller, W.E.L. Grimson and W.M. Wells. *Efficient Population Registration of 3D Data*. In Computer Vision for Biomedical Image Applications, pages 291–301, 2005. 116